U.PORTO

FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

**MIEEC - MESTRADO INTEGRADO EM ENGENHARIA ELETROTÉCNICA E DE COMPUTADORES** 2013/2014

A Dissertação intitulada

*"Very Short-Term Solar Power Forecast for MV/LV Distribution Grids"*

foi aprovada em provas realizadas em 09-04-2014

o júri

Presidente **Professor Doutor Manuel António Cerqueira da Costa Matos**
Professor Catedrático do Departamento de Engenharia Eletrotécnica e de Computadores da Faculdade de Engenharia da Universidade do Porto

**Professor Doutor Jean Sumaili**
Professor Auxiliar Convidado da Universidade Lusófona

**Professor Doutor Vladimiro Henrique Barrosa Pinto de Miranda**
Professor Catedrático do Departamento de Engenharia Eletrotécnica e de Computadores da Faculdade de Engenharia da Universidade do Porto

**Doutor Ricardo Jorge Bessa**
Investigador do INESC - TEC

O autor declara que a presente dissertação (ou relatório de projeto) é da sua exclusiva autoria e foi escrita sem qualquer apoio externo não explicitamente autorizado. Os resultados, ideias, parágrafos, ou outros extratos tomados de ou inspirados em trabalhos de outros autores, e demais referências bibliográficas usadas, são corretamente citados.

**Autor - Artur Jorge Teixeira Trindade**

Faculdade de Engenharia da Universidade do Porto

Faculdade de Engenharia da Universidade do Porto


# Very Short-term Solar Power Forecast for MV/LV Distribution Grids

Artur Jorge Teixeira Trindade

FINAL VERSION

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores
Major Energia

Supervisor: Pr. Dr. Vladimiro Henrique Barrosa Pinto de Miranda
Second Supervisor: Dr. Ricardo Jorge Gomes Sousa Bento Bessa

March, 2014

# Abstract

The growing penetration of dispersed energy source in electric networks poses new challenges for operators in the distribution network, particularly related to voltage control and power flow direction. Moreover, the electricity markets liberalization has enabled the emergence of the "consumer-producer", opening doors to new tariffs, far more flexible.

To promote the use of energy from renewable sources, smart grids, that are already a reality, should play a key role in the future of electric systems worldwide.

Given the contribution of photovoltaic (PV) source in dispersed generation and its variability, it is essential that Distributed System Operators (DSO) and electricity market agents have at their disposal PV generation forecasting tools as accurate as possible.

To address this need several methods have been developed for PV forecast. The behavior of clouds is critical to PV production, thus, to introduce information about cloud presence in forecasting system, there are two families of methods: methods based on Numerical Weather Predictions (NWP) and methods based on satellite imaging.

This work is part of a third approach, on which to our knowledge there are only two previous works. In this thesis, we propose a method that uses PV generation data of the neighborhood to provide information on the cloud impact in the generation of a regional profile system. These methods may be called spatial-temporal.

In this work are used the properties of the vector autoregressive (VAR) models to introduce spatial information in the forecasting system. Then the results are compared with the auto-regressive models, which do not allow the introduction of spatial information.

In this work we propose three important innovations:

• The use of data from a lower level, data from energy box (EB) in the forecast generation at a higher level of network, the secondary station MT/BT level, where is installed the Distribution Transformer Controller (DTC).

• The development of adaptive models in time, based on the recursive least squares method

• The development of a non-adaptive model, which can deal with high dimension problems, through variable selection, the Gradient Boosting univariate and multivariate analysis.

The results show that the introduction of the EB information in the DTC prediction brings about an improvement of about 4% in the non-adaptive, for models that do not use this information models, allowing predictions nRMSE 8.8%, whereas the adaptive models that improvement is 5.1% to 8.4% of predictions nRMSE.

The introduction of spatial and temporal information improves the overall performance by 16% in some cases enabling improvements exceeding 45% in non-adaptive models, while the adaptive models present an overall improvement to 15.4% surpassing the 25% in isolated cases.

Key words: Smart Grids, Photovoltaic forecasting, Spacio-temporal methods, Distribution System Operator, Electricity markets agents

# Resumo

A tendência crescente para a introdução de energia de origem dispersa nas redes elétricas coloca novos desafios aos operadores da rede de distribuição (DSO), nomeadamente relacionados com o controlo de tensão e sentido dos trânsitos de potência. Por outro lado, a liberalização dos mercados de eletricidade possibilitou o aparecimento do produtor-consumidor, abrindo portas a novas tarifas, bastante mais flexíveis.

Para potenciar a utilização de energia de fonte renovável, As redes inteligentes, sendo já uma realidade, devem desempenhar um papel fundamental no futuro dos sistemas elétricos de todo o mundo.

Atendendo ao papel da energia fotovoltaica na produção dispersa e à sua variabilidade, é fundamental que tanto os DSO como os agentes de mercado disponham de ferramentas de previsão de geração fotovoltaica, tão exatas quanto possível.

Para dar resposta a esta necessidade, vários têm sido os métodos de previsão fotovoltaicos desenvolvidos. O comportamento das nuvens é crítico para a produção fotovoltaica, assim no sentido de introduzir no sistema de previsão informação sobre a presença de nuvens existem duas famílias de métodos: os métodos baseados em previsões meteorológicas e os métodos baseados em imagens de satélite.

Este trabalho insere-se numa terceira abordagem, sobre a qual apenas são conhecidos dois trabalhos anteriores. Nesta dissertação propõem-se um método que utiliza informação da geração fotovoltaica da vizinhança para fornecer ao sistema de informação sobre o impacto das nuvens no perfil de geração de uma região. Estes métodos poderão ser denominados de espácio-temporais.

Neste trabalho utilizam-se as propriedades dos modelos autorregressivos vetoriais para introduzir a informação espacial e comparam-se os resultados com os modelos autorregressivos, que não possibilitam a introdução de informação espacial.

Neste trabalho propõe-se 3 inovações importantes:

A utilização de dados de um nível inferior, dados provenientes de contadores inteligentes *("Energy Box",* EB) na previsão de geração a um nível superior da rede, ao nível do posto de transformação MT/BT, onde está colocado o "*Distribution Transformer Controller*" (DTC).

O desenvolvimento de modelos adaptativos no tempo, baseados no método recursivo dos mínimos quadrados

O desenvolvimento de um modelo não adaptativo capaz de lidar com problemas de grande dimensão através de seleção de variáveis, o Gradient Boosting univariado e multivariado.

Os resultados mostram que a introdução da informação das EB na previsão do DTC introduz melhorias de cerca de 4% nos modelos não adaptativos, relativamente aos modelos que não utilizam essa informação, permitindo previsões com nRMSE de 8.8%, ao passo que nos modelos adaptativos essa melhoria é de 5.1% atingindo previsões com 8.4% de nRMSE.

A introdução de informação espácio-temporal melhora os resultados globais até 6% possibilitando em alguns casos melhorias superiores a 45%, nos modelos não adaptativos, enquanto os modelos adaptativos multivariados apresentam melhorias globais até 15.4% ultrapassando em casos isolados os 25%.

Palavras-Chave: Redes Inteligentes, Previsão fotovoltaica, Métodos espácio-temporais, Operadores redes de distribuição, Agentes mercados eletricidade

# Acknowledgements

Firstly, I would like to thank my Supervisors at INESC TEC, Pr. Vladimiro Miranda and Dr. Ricardo Jorge Bessa, for their unwavering support, motivation, guidance and expertise that helped the completion this work and their contributions to my graduation and my identity as a future Electrical Engineer.

My sincere gratitude towards my teachers who helped in my graduation, who taught me more than knowledge about this course and whose expertise and their way of being in life I will always admire.

I would also like to thank my colleagues, who shared with me the open space at INESCTEC, for their support.

My special thanks go to Hélder Franco and David Otero for all their support and suggestions thoroughout these last months, and off course to my best man José Carlos Santos for all his support on last ten years.

My special thanks to my father my mother and my syster for raising and educating me and their ever present companionship and support. There are no amount of words that I could use to thank the person who supported me the most and gave me the strength and confidence to face this last challenge in my graduation, my love Silvia Carvalho.

Finally, I thank every person who helped shape me into the person I am now, all of whom are still with me, those who unfortunately are not and those whose work and philosophies influenced me the most.

This work was made in the framework of the BEST CASE project ("NORTE-07-0124-FEDER-000056") financed by the North Portugal Regional Operational Programme (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF), and by national funds, through Fundação para a Ciência e a Tecnologia (FCT). It was also developed in the framework of the FCT projects «SMAGIS – PTDC/SEN-ENR/113094/2009» and «DYMONDS – CMU-PT/SIA/0043/2009», co-financed by the COMPETE Programme.




# Contents

# List of Figures

# List of Tables

# Abbreviations and Symbols

**Abbreviations List**

| | |
|---|---|
| ACF | AutoCorrelation Function |
| ADSL | Asymmetric Digital Subscriber Line |
| AFSOL | Aerosol-based Forecast of SOLar irradiance |
| AIC | Akaike Information Criterion |
| ANFIS | Adaptive Neuro Fuzzy Inference System |
| ANN | Artificial Neural Network |
| AR | Autoregressive |
| ARIMA | AutoRegressive Integrated Moving Average |
| ARPS | Advanced multiscale Regional Prediction System |
| ARX | AutoRegressive with eXogenous input |
| BEM | Mean Bias Error |
| CV | Cross Validation |
| DSO | Distribution System Operator |
| DTC | Distribution Transformer Controller |
| EB | Energy Box |
| ECMWF | European Center for Medium-range Weather Forecasts |
| EURAD | EURopean Air pollution Dispersion |
| GA | Genetic Algorithm |
| GB | Gradient Boosting |
| GB multi | Gradient Boosting multivariate |
| GB uni | Gradient Boosting univariate |
| GEM | Global Environmental Multiscale |
| GFS | Global Forecast Systems |
| GHI | Global Horizontal Irradiance |
| GNWP | Global Numerical Weather Predictions |
| GPRS | General Packet Radio Service |
| GVP | Grid-Point Value |
| GW | Giga Watt |
| Imp | Improvement |
| kNN | k-Nearest Neighbors |
| LAN | Local Area Network |
| LV | Low Voltage |
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percentage Error |
| MLP | MultiLayer Perceptron |
| MM5 | Mesoscale Model MM5 |
| MOS | Model Output Statistics |
| MV | Medium Voltage |
| MVF | Motion Vector Field |

| | |
|---|---|
| NDFD | National Forecast Data Base |
| NMAE | Normalized Mean Absolute Error |
| NRMSE | Normalized Root Mean Square Error |
| NWP | Numerical Weather Predictions |
| NWS | National Weather Service's |
| OLS | Ordinary Least Squares |
| PLC | Power Line Carrier |
| PV | Photovoltaic |
| QoS | Quality of Service |
| RBF | Radial Basis Function |
| RBFN | Radial Basis Function Network |
| RLS | Recursive Least Squares |
| RMSE | Root Mean Square Error |
| rRMSE | relative Root Mean Square Error |
| SbF | Satellite-based Forecasts |
| SOM | Self-Organizing Maps |
| STPP | Solar-Thermal Power Plants |
| SVM | Support Vector Machines |
| TDNN | Time Delay artificial Neural Network |
| UMTS | Universal Mobile Telecommunications System |
| VAR | Vectorial AutoRegressive |
| VARX | Vectorial AutoRegressive with eXogenous input |
| WAN | Wide Area Network |
| wqr | weighted quantile regression |
| WRF | Weather Research and Forecasting model |

**Symbols List**

| | |
|---|---|
| $\hat{p}_{t+k}$ | forecasted solar power k-hours-ahead from prediction time |
| $\Gamma$ | correlation matrix |
| $doy$ | day of the year |
| $\lambda$ | forgetting factor |
| $p_{t+k}$ | measured solar power k-hours-ahead from prediction time |
| $p_{t-l}$ | measured solar power l-hours-delayed from prediction time |
| $ng$ | number of groups used in Cross Validation function |
| $mmax$ | number of iterations used in Cross Validation function |
| $mstop$ | number of iterations used in Gradient Boosting function |
| $P_{inst}$ | Photovoltaic installed power |
| $\tau$ | quantile probability |
| $U$ | residuals |
| $\nu$ | shrinkage factor |
| $sdoy$ | size of the kernel for the day |
| $sh$ | size of the kernel for the hour |
| $\hat{p}_t^{cs}$ | solar power production estimation on a clear sky day |

# Chapter 1

# Introduction

## 1.1 Motivation – The importance of photovoltaic forecasting

The smart grids, sustained in dispersed generation, the emergence of consumer - producers, the development of new telecommunication technologies/protocols and new tools of computational intelligence, exist today and should play a decisive role in the future.

Changing the concept of distribution networks unidirectional to bidirectional poses new challenges to the Distributed System Operator (DSO), such as voltage control, grid protection, flickers, frequency volatility and fault levels.

In this context, distribution networks fail to be unidirectional networks, where energy flows from the substation to the point of use, becoming bidirectional where power flow can occur in both directions. Thus, not only the correct prediction of the load, but more recently, the micro-generation acquires fundamental importance.

Due to his rooftop installation, PV production takes an important role in these bidirectional grids.

Photovoltaic (PV) technology is playing an important role in disperse generation growing. The rooftop PV segment, in Europe, in 2012, was nearly fifteen gigawatt (GW).Thus PV distributed generation will have a key role in both electric system operation and electricity markets.

Although 2012 was a year of economic crisis, the capacity installed, which has been estimated in 31 GW new PV all over the world, has almost reached the record-setting year of 2011 [1].

In 2012, the cumulative generation of electrical power installed using PV technology surpassed the 100 GW reaching 102GW. This cumulative power capability is able to produce, every year, the equivalent energy produced by 16 coal power plants or nuclear reactors of 1 GW each, saving fifty three million tons of $CO_2$.

After being for the second consecutive year the number-one new source of electricity generation installed in Europe, PV technology covered in 2012 2.6% of the demanded electricity and 5.2% of the electricity peak.

# 2 Introduction

While in Europe PV is already an important player in the electricity mix, the expansion of this technology is expected to accelerate on the rest of the world, in the next years.

A PV increase of competitiveness explains this behavior. The general rise in energy prices (mainly fossil) and the strong price decreases in PV technology, falling faster than expected, led the market to the "grid parity".

"Grid parity" can be explained as the leveled cost of generate electricity, which equals the cost of electricity grid, i.e., production and transportation costs. "Grid parity" point is important because once achieved this level the technology, here referred to PV power production, became economically competitive without subsidies, i.e., it is more independent of political decisions.

As result of this evolution, PV technology is increasing its penetration in electricity system.

The changes produced in electrical systems by distributed PV generation, as other micro-generation technologies, led to an emergence of the consumer/micro-producer.

Smart Grids provide the necessary tools to efficient operation of the electrical system, based on the new paradigm presented above. Évora Inov Grid is able to monitor, analyze, process, store, report and act at distributed level, across multiple sensors, actuators and devices, placed on all relevant points of the network, and communicating among themselves and with all other stakeholders [2].

With the advent of smarts grids new customers for new services are emerging. It is expected that consumer/micro-producer will be a much more active player in market than the classic consumer.

To maximize renewable energies penetration, without compromising the system security, Distributed System Operator (DSO) needs good forecasting tools, to know the relation between generation and consumption to keep voltage levels in technical range. While, electricity market agents, which in this new paradigm could be, theoretically, every consumer, need good forecasting tools, in order to reach maximum benefits.

Very short-term forecasting became crucial for electricity market agents because that the more volatile the relationship between generation and loads, more volatile energy prices should be. The volatility of energy prices throughout the day open new market opportunities.

In recent years there have been several developments in this area. The knowledge of the behavior of photovoltaic generation for relatively short time horizons has great importance for network operators (either transport or distribution) of the electrical system. It also contributes to using energy more efficiently leading to economic gains, especially as the cost of PV energy produced locally approximates the tariff parity.

Photovoltaic forecasting has important dependence on cloud behavior. In order to provide the system cloud behavior information methods normally use Numerical Weather Predictions (NWP) or satellite images.

This dissertation proposes a different forecasting approach based on spacio-temporal methods, which, we are aware, there are only two works using these methods in photovoltaic forecast. Using temporal information of neighborhood, i.e. generation data series from neighbor sites, to provide the system cloud behavior geographical information. Since the PV generation of neighbor sites is correlated with the cloud presence, Vectorial Autoregressive (VAR) methods introduce cloud information to the forecasting system.

## 1.2 Objectives of this Dissertation

In this Dissertation a new very short-term forecasting concept it is proposed. Aiming to emulate the cloud behavior, as alternative to numerical weather predictions (NWP) and satellite images, develop Vector Autoregressive (VAR) models that use as input information from neighboring.

Two families of models, based on Classic VAR method, can be found in this dissertation. VAR models relate all the input variables with all response structures, which does not happen in autoregressive (AR) traditional methods. The use of VAR models implicitly introduces cloud information in forecasting system. Aiming to deal with possible changes in photovoltaic (PV) panel conditions, and in their neighborhoods affecting PV electrical generation, proposes the Recursive Least Squares (RLS), as on-line adaptive VAR model. Gradient Boosting (GB) off-line, non-adaptive VAR models are developed aiming to handle with the dimension problem, which can rapidly increase in VAR, and select the most important information for each site.

In both methodologies, it is tested the use of recorded PV generation values, from a lower level as exogenous input variables, referred as VAR with exogenous VARX, which as far as we know is also an innovative approach.

The new contributions of this dissertation to PV distributed forecasting area are the use of VAR models to provide cloud behavior information to the forecasting system, avoiding the necessity of NWP or Satellite images as input variables, i.e., using spatial-temporal methods aiming to achieve a more accurate prediction.

Also new, is the use of Gradient Boosting (GB) approach, which should be able to deal with the high dimensional problem selecting weighted input variable coefficients according to the information provided to the system in a more effective way than the classical VAR models.

## 1.3 Structure of this Dissertation

Several works and papers were reviewed aiming to better understand PV forecasting science and realize what has been done in this area of knowledge. Chapter 2 provides a literature review of the most important PV forecasting methodologies as well as some results and considerations presented by the authors.

Chapter 3 features the main concepts and the tools that were developed and adapted to the PV forecasting problem. Initially, due to VAR models necessity of centralized computation, and to explain how raw data becomes available, section 3.1 presents the InovCity smart grid communication infrastructure. Following, due to PV data series seasonality, both in day and year range, and attending that autoregressive (AR) and VAR classical models assume data series stationarity, section 3.2 presents a clear sky model to provide stationarity to data. The methodologies proposed here are based on classical AR and VAR models, thus sections 3.3 and 3.4 devote to explain classic AR and VAR models, respectively. Spatial-temporal models section divided in VAR and VARX. Subsequently, section 3.5, describes the Recursive Least Squares (RLS) based adaptive methodology. Afterwards, sub sections 3.6.1 and 3.6.2 explain the Gradient Boosting (GB), univariate and multivariate version, basis and their application to the problem of this dissertation. Finally, in section 3.6, sub section 3.6.3 presents the cross

validation algorithm, which calculates the optimal number of iterations to use in GB providing the better results.

To understand if this approach is useful, several tests and analysis are done. However, before results and their analysis, Chapter 4 provides a brief presentation of the case study.

Finally, the main conclusions taken from the results analysis and suggestions for further works are presented in Chapter 5.

# Chapter 2

# State of the art

This chapter provides an overview on both solar power and solar irradiation forecasting which have been usually classified in three kinds of models: Physical models (section 2.1), Satellite based models (section 2.2) and Statistical models (section 2.3).

Regarding the time-scale of the predictions, the statistical models have been separated into three types: i) very short-term forecasting (section 2.3.2) which includes forecasting models that make predictions until six hours-ahead; ii) short-term forecasting (section 2.3.2), from 6 hours up to one week-ahead; iii) and lastly spatial-temporal models (section 2.3.3).

During the last decades, the demand of energy all around the world has been increased. Especial attention has been focused on the use of renewable energy sources. For this purpose, a better understanding of forecasting both photovoltaic (PV) and wind power is required. As result, several studies have been performed in both solar power and solar irradiation forecast [3].

Solar irradiance depends on seasonal and daily variation but also on the climatic behavior. Distinguishing between both kinds of variability becomes a critical issue on PV forecasting. One can divide the models according to how they handle the variability of irradiance. In physical models, Numerical Weather Predictions (NWP) were included; Satellite-based Forecasts (SbF) employ satellite images in modeling; and in Statistical models, inputs, such as past values of solar irradiance or solar power production, were used.

Due to the fact that larger forecast errors increase financial losses, root mean square error (RMSE) is commonly used since this metric weights in larger deviations. Other metrics are also used in model accuracy assessments such as mean absolute error (MAE) or bias. To emphasize the error results it is common use error values in percentage. Thus, a normalization is required. Several normalizations are used, but normally the same name is used. In this chapter will be presented the normalizing equation near to the results which is referred.

A naïve model often used in forecasting works is the Persistence model. Persistence consists in take as prediction the last known value of the variable to be forecasted, without any computational work.

The chapter is concluded with final remarks and the original contributions which of this dissertation.

## 2.1 Physical models

In order to predict solar power production, models including analytical function of the PV panels behavior are designated as Physical models. These models are based on NWP and they have been reviewed in this section.

Due to scientific and computational development, weather prediction models became more complex and global leading to models composed by three main components: “dynamical”, physical and data assimilation. As the models accuracy became better, the time horizons forecast became greater. Therefore, both physical parameterization and the development of statistical tools contributed to more effective and reliable models [3].

Despite of their developments, models still need the input of initial state conditions, which are determined with measurement-errors, and boundary conditions that cannot be strictly defined.

In order to operate a local NWP model, the knowledge of several earthly parameters (e.g., land-sea masks, soil data, and topography) is required and it is obtained just one time for each location. Moreover, it is necessary to download global predictions, more than once per-day from low-resolution models (typically 50 – 100km horizontal grid sizes), and to include in the model local data for the forecasting sites.

Numerical Weather Predictions (NWP) are usually divided in Global models and Mesoscale models. The latter ones provide regional scale forecasting with resolution ranging from 2 km to 50 km [4]. Although Global models show lower spatial-resolution (200 km), they are able to provide predictions up to one week-ahead in time. Global and Mesoscale NWP are summarized in Table 2.1 and Table 2.2, respectively.

Table 2.1 – Global NWP, adapted from [4]

| **Model** | **Developer** |
|---|---|
| GFS – Global Forecast System | National Centre for Environmental Prediction |
| GEM – Global Environment Multiscale Model | MSC – Meteorological 00Service of Canada |
| GME - Operational Global Icosahedral-Hexagonal Gridpoint Model | German Weather Services |
| IGCM – Intermediate General Circulation Model | Department of Meteorological at the University of Reading |
| UM – Unified Model | UK Met Office |

Global Forecast Systems (GFS) is a 1° x 1° resolution global meteorological variables forecasting system such as temperature or radiation among others. These systems use measured weather in order to obtain atmospheric behavior from both dynamic and thermodynamic points of view, and to solve an equation model. Mesoscales are regional weather forecast models that use, as inputs, planetary models outputs and boundary conditions. Global models have low-

resolution and it is needed to work on their outputs in order to achieve reliable local predictions [5].

Table 2.2 – Mesoscale NWP, adapted from [4]

| Model | Developer |
|---|---|
| ALDIN – Aire Limitée Adaptation Dynamique Développment InterNational | Météo-France |
| ARPS – Advanced Regional Prediction System | University of Oklahoma |
| GEM-LAM – Global Environmental Multiscale Limites Area Model | MSC - National Laboratory for Marine and Coastal Meteorology |
| MM5 – Fifth Generation Penn State/ NCAR Mesoscale Model MM5 | PSU/NCAR - Pennsylvania State University / National Center for Atmospheric Research |
| WRF – Weather Research and Forecasting Model | NCEP - National Centers for Environmental Prediction |

A Mesoscale weather prediction, like the Fifth Generation Penn State/ NCAR Mesoscale Model MM5 (MM5), can be used with the purpose of improving model's resolution. MM5 was used to construct a 1km x 1km resolution NWP model. In fact it is possible to use a model with less resolution (3km x 3km) given that this model present almost the same predictions but the computational cost is smaller.

Due to their low-resolution, global numeric weather predictions (GNWP) are not suitable for local whether predictions [6]. Although every approaches start with GNWP, such as European Center for Medium-range Weather Forecasts (ECMWF) or GFS, the subsequ0ent step to achieve a local weather forecast requires the support of a Mesoscale NWP such as the Weather Research and Forecasting model (WRF), the statistical processing as Model Output Statistics (MOS) or hybrid models combining both techniques. Based on satellite information and on historical values, ECMWF-OL models were used to analyze irradiation in Germany. These models presented the best results referring to RMSE for all three time horizons which are one, two and three days ahead and obtaining 40.3%, 41.6% and 44.9%, respectively, while Persistence achieved was 63.5%, 70.2% and 73.3%. The same method has achieved the best results in both Swiss and Spanish cases. Regarding Austrian irradiation, the best values were presented by statistical Blue Sky model but these can be applied only in Austria.

Although Mesoscale based models present worst results than global models, where a post-processing is required, these models present a positive improvement compared to the Persistence model.

*Remund et al.* [7] compared three NWP models applied in three distinct USA climates for few hours-ahead, one day-ahead and two days-ahead irradiance forecasting: ECMWF, National Forecast Data Base (NDFD) and GFS/WRF. Although the three models present a very stable RMSE, Persistence error increases with time horizon and the best results show that RMSE varies from 30% for one hour-ahead to more than 60% for two days-ahead forecast. While ECMWF and NDFD models present a RMSE around 40% for all time horizons, GFS/WRF presents a value slightly below 50% for the time horizons lower than one day and 50% for one and two days-ahead predictions.

Global Environmental Multiscale (GEM), ECMWF, two WRF versions, a Mesoscale model designated by MASS, the Advanced multiscale Regional Prediction System (ARPS) and a cloud

cover based forecast model were compared by *Perez et al.*[8]. Both GEM and ECMWF are global models, while the remaining are Mesoscale models. Authors concluded that GEM and ECMWF presented the best performance.

*Zamora et al.*[9] evaluated the Eta Model, which is a Mesoscale physical model that forecasts hourly solar irradiance using solar hour mean zenith angle as input. Authors found that solar irradiation at ground level showed dependence on both cloud presence and atmospheric aerosol levels.

Concerning the relation of NWP and measures, *Bacher et al.*[10] presented shadow as an explanation for systematical error. Using a statistical based clear sky model, the authors were able to correct the shadow effect. The advantage of this method is that it is not necessary to install sensors in or in the vicinity.

*Ohtake et al.* [11] studied the accuracy of solar forecast using a model based on grid-point value (GVP) dataset of Mesoscale models and they concluded that global horizontal irradiance (GHI) is generally underestimated in summer and overestimated in winter.

In order to optimize the use of Solar-Thermal Power Plants (STPP), *Wittman et al.* [12] developed an aerosol-based forecast of solar irradiance for energy applications (AFSOL). It includes sub-models for the treatment of the chemical transport model, emission data and meteorological input. In this case, the European Air Pollution Dispersion (EURAD) model provides information about aerosol, water vapor and cloud coverage data. For clear sky days, while the AFSOL forecast overestimates nearly 3.9% of the measured values, ECMWF results underestimate by 11.9%. In terms of RMSE, AFSOL is more accurate than ECMWF with 5.2% and 12.4% RMSE respectively. However, ECMWF became better with RMSE value of 18.5% against 25.1% obtained for AFSOL, when all days of the test period are considered.

## 2.2 Satellite-based Forecasts

As an alternative to NWP and aiming to get relevant information of solar behavior, satellite-based forecasting models have been suggested as feasible candidates. They use recent clouds motion in order to forecast the clouds location in a near future.

Cloud presence has huge influence on irradiance levels. METEOSAT provides good temporal and spatial resolution information. Due to the lack of measure sensors at ground level, geostationary satellites METEOSAT-image based models have been used successfully, as solar irradiance sensors. In fact, the results obtained from the satellite-image based models have shown as good as the measurements of a ground station at a distance of 25 km". The HELIOSAT model, based on physical parameters, is an example of the use of satellite images to calculate ground level solar irradiance. Based on this knowledge [13] described an image forecasting model with the goal of getting up to two hours-ahead solar irradiance forecasting model.

Taking two consecutive cloud index images, a statistical method was applied in order to get smoothed images. Assuming constant pixels intensities, constant gradients and smoothness of vector field, a Motion Vector Field (MVF) was determined. When MVF to an image a new image is obtained, as a forecasting image. Then, HELIOSAT model is applied to the forecasted image providing the solar irradiance forecast value.

Although it should be noted that the magnitude of smoothing increases with the time horizon, the results have shown that the satellite images based model delivers better results than the Persistence model for all studied time horizons (up two hours-ahead).

*Ahlstrom and Kankiewicz* [14] also generated a satellite image based forecast model. By extrapolating the information of two consecutive images with 30 minutes lag between them i.e. an image at the moment $t_0$ and another one obtained thirty minutes earlier ($t_0$-30 minutes), the model generates a Motion Vector Field (MVF) which forecast the ($t_0$+30 minutes) image.

The authors compared four forecasting models, such as the classical Persistence, a NWP and two models based on image extrapolation, up to 6 hours-ahead. The difference between the two models, based on image extrapolation, was that a statistical smooth version of MVF was or was not used, respectively.

It was suggested that the smoothed MVF model performs better than Persistence and MVF without smoothing for every time horizon. On the other hand, NWP presents lower RMSE than Persistence from two hours-ahead. Therefore, while MVF with smoothing showed to be the best model until five hours-ahead, the NWP based became the best model after five hours.

Also, NWP showed better results than MVF without smoothing from nearly above three hours-ahead.

## 2.3 Statistical Models

Normally the statistical models have a two-stage approach. First, the irradiance or solar power is normalized using a clear sky model in order to eliminate seasonal and daily variation and achieve a more stationary time series. Clear sky models can be analytical, based on solar relative position, or statistical, using statistical smoothing techniques. Then, time series or machine learning methods can be applied to forecast future values of time series.

### 2.3.1 Very short-term

According to the literature, the NWP model (described in section 2.1) only improves solar power forecasts in time horizons larger than four hours. The Persistence model is hardly beaten for minutes-ahead forecasting and it is also a competitive model for very short-term, namely one to two hours-ahead. For few hours-ahead both satellite-based models and NWP models are mainly used [15].

An online short-term solar power forecasting with a two stage approach is described by *Bacher et al*. [16]. Firstly, solar generation is normalized with a clear sky model based on statistical smoothing techniques and quantile regression achieving a more stationary time series. Then, a linear model is used to generate power forecast. These models are adaptive and they are able to consider variations in PV production, due to dirt on panel or snow cover among others.

Three different linear models were compared:

- An autoregressive (AR) model using lagged measures as input.
- A model designated as $LM_{NWP}$ which is a linear model where a normalized NWP were used as input.
- Autoregressive with exogenous input (ARX), which is the NWP.

Using the autocorrelation function (ACF), authors have defined the most important variables to consider in the models.

Taking RMSE as reference (calculated as in Eq. (2.1), the authors have improved the results by around 35% when the ARX model was used instead of the reference model in the first six hours-ahead forecast. While in very short-term, up two hours-ahead, the known values of the

last few hours became the most important input, for longer time horizon NWP showed to be the main input. In fact, the ARX model is slightly better than AR for the first six hours. This indicates that although the introduction of NWP as input has a positive impact for all time horizons, the importance of NWP increases with time horizon.

$$RMSE_k = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(p_{t+k} - \hat{p}_{t+k})^2} \tag{2.1}$$

where:
$p_{t+k}$ is the measured solar power k-hours-ahead from prediction time
$\hat{p}_{t+k}$ is the forecasted solar power k-hours-ahead from prediction time
*N* is the prediction sample size

*Pedro and Coimbra* [15] compared different models for very short-term forecasting. In this analysis, past values of the time series were used as input. The studied models were: Persistence, autoregressive integrated moving average (ARIMA), k-nearest neighbors (kNN), artificial neuronal networks (ANN) and hybrid genetic algorithm/artificial neuronal networks (GA/ANN), the genetic algorithm was used to optimize ANN architecture.

A clear sky approximation based not on physical parameters but in statistical smooth technique function was applied to every listed models, including Persistence.

Considering several statistics metrics, such as mean absolute error (MAE) (described in Eq. (2.2), mean bias error (MBE) (Eq. (2.3), and normalized root mean square error (NRMSE) (Eq. (2.4), the test data was divided in 3 periods known as high, medium and low variability [17].

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|p_t - \hat{p}_t| \tag{2.2}$$

$$MBE = \frac{1}{N}\sum_{t=1}^{N}p_t - \hat{p}_t \tag{2.3}$$

$$NRMSE = \sqrt{\frac{1}{N}\frac{\sum_{t=1}^{N}(p_t - \hat{p}_t)^2}{\sum_{t=1}^{N}\mathrm{p_t^2}}} \tag{2.4}$$

Although GA/ANN generally provides the best results in average, the kNN model presents both the best and the worst results. For the kNN model, the NRMSE is 6.38% for low variability period (July 1st, 2011 to August 15th, 2011) in one hour-ahead forecast, and 37.86% for the high variability period (January 1st, 2011 to April 30th, 2011) in two hours-ahead forecast. While for GA/ANN model NRMSE varies between 11.42% for middle variability period (May 1st, 2011 to June 30th, 2011) in one hour-ahead forecast, and 24.39% for low variability period in two hours-ahead forecast.

*Mellit and Pavan* [18] proposed a multilayer perceptron (MLP) ANN for two hours-ahead forecast of a grid connected PV plant in Triest. The results showed that the forecasting model behaves well for sunny days, presenting a MBE between 7.33% and 32.00%, a correlation coefficient above 98.95% and a RMSE between 18.98% and 67.08%. However, its performance

decreased in cloudy days, when the MBE ranges between 45.55% and 53.67%, correlation coefficient varies from 92% to 95% and RMSE interval is between 54.67% and 85.76%.

*Huang et al.*[19] used a Fourier time series based model to remove the seasonal component of solar irradiance. A second order autoregressive model AR(2) as well as a Lucheroni's and a combination model were tested and compared aiming to get a one hour-ahead forecasting model. While in the AR(2) model, the solar radiation decreases make that error also decreases, the Lucheroni's model has performed very well in peak prediction. In order to use the characteristics of both models, a combination model was developed which is based on a difference equation. When a one hour-ahead forecast is required, the differences between the current moment and the forecast obtained 1 and 2 hours earlier should be analyzed. If both differences are negative, the AR(2) model must be considered. Otherwise Lucheroni's should be used.

The results were: AR(2) presents the smallest NRMSE (18,59%) (calculated as in Eq. (2.5)); combination of models achieve almost the same NRMSE (18,92%); Lucheroni's model obtain a NRMSE equal to 19,09%.

$$NRMSE = \frac{\sqrt{\frac{\sum_{t=1}^{N}(\hat{p}_t - p_t)^2}{N}}}{\frac{1}{N}\sum_{t=1}^{N} p_t} \tag{2.5}$$

*Wu and Chan* [20] worked on 10 min-ahead solar irradiation prediction comparing time delay ANN (TDNN), ARMA and an Hybrid model. While ARMA succeeded in understanding linear problems, TDNN was adequate in studies with non-linear complexity. Especially in which outliers concerns TDNN outperforms ARMA. So, and in order to lead with both linear and non-linear problems, a two stage hybrid model was constructed. Firstly, the ARMA model is used to deal with linear data component. Then, the non-linear behavior is captured by TDNN.

In order to deal with the non-stationary solar radiation issue, the authors applied four analytical different methods: the Jain's model based on exponential function; the Baig's model which is an evolution of the Jain's model including both an exponential term and a cosine function; the S. Kaplanis' model cosine function based; and a second order polynomial function, AL-Sadah's model. Then, the AL-Sadah's model was choose because it shown both the better detrending and the best fit.

Authors concluded that ARMA model was generally able to predict the solar series behavior. However, the main problem was that it showed a lag. On the other hand, TDNN did not show an obvious lag but it was not capable of capture the peak. Therefore, the hybrid method was suggested because it showed the best results in a 10 months forecasting window (January, 2010 – October, 2010).

### 2.3.2 Short-term

While in very short-term forecasting those models without NWP can achieve good results, on short-term horizons NWP became essential to obtain a forecast with acceptable accuracy.

*Bacher et al.* [16] used NWP as exogenous variables in a forecasting model (described already in section **Error! Reference source not found.**). NWP were the most important orecasting inputs for time-horizons longer than two hours. For time horizons between 19 and 29 hours-ahead, both $LM_{NWP}$ and ARX have presented similar behaviors. Their analysis were

better than that obtained by the AR model indicating that past values did not show a significant influence in forecasting for this time horizons.

Using NWP as input, *Jimenez et al.*[5] compared several possible models such as Persistence, ARIMA, kNN, ANN and adaptive neuro-fuzzy (ANFIS).

In this work, the classical Persistence model was not utilized. Instead of using the forecast value, obtained during the last hour, the authors used a model in which the forecasted values between 09:00 and 00:00 was obtained from measuring the previous day.

In addition to NWP, past values of hourly electrical energy generation were used as input in model construction. NWP were produced by MM5.

In terms of NRMSE (calculated with Eq. (2.6) MLP ANN showed an average error of 11.79% for the time horizon between 16 and 39 hours.

$$NRMSE = \frac{RMSE}{P_{inst}} \tag{2.6}$$

*Marquez et Coimbra* [21] used a US National Weather Service's (NWS) report as an input for an ANN model. In combination with Genetic Algorithm (GA), the optimal subset of inputs was optimized. The main goal was to develop and validate several "medium-term solar forecasting". The ANN was trained with 11 input variables, divided in two "geometric/temporal" variables (related with clear sky diurnal variations) and 9 meteorological forecasts.

In terms of rRMSE (calculated with Eq. (2.7) the best model achieved an error of 17.7% when the sky cover (%), probability of precipitation (%), minimum temperature and "cosine of the solar zenith angle" were included as inputs. On the other hand, Persistence presented an error of 31.1%.

$$rRMSE = \frac{RMSE}{\frac{1}{N}\sum_{t=1}^{N}\hat{p}_t} \tag{2.7}$$

where $\hat{p}_t$ is the forecasted value and $N$ is the sample size.

*Paoli et al.*[22] also used ANN for irradiation forecast. They identified ARIMA (classic method), Bayesian inferences, Markov chains and kNN as the most the popular solar forecasting methods and they developed an ANN model which was compared with Persistence, Markov chain (order 3), Bayes (order 3), kNN (order 10) and AR (order 8). The AR and ANN results were compared using both non-stationary and stationary time series as input. In order to get the stationary time series both clear sky and clearness indexes were used. Table 2.3 shows a summary of the results.

Table 2.3 – Results obtained in[22]

| | **RMSE%** | | |
|---|---|---|---|
| | **non-stationary** | **Clearness index** | **Clear sky index** |
| **Persistence** | **26.13** | **-** | **-** |
| **AR (order 8) without pre-processing** | **21.18** | **-** | **-** |
| **ARMA (2,2) with pre-processing** | **-** | **20.31** | **20.32** |

| | | | |
|---|---|---|---|
| **ANN** | **20.97** | **20.17** | **20.25** |

In order to get a three steps irradiance forecasting model with three days range, *Lorenz et al*. [23] used ECMWF low resolution NWP and a "spatial averaging and temporal interpolation" model was developed.

Since distance dependent weights did not improve the results, the arithmetic average of surrounding pixels was proposed. Then, a titled irradiance model converts global horizontal irradiance. Finally, with the aim to correct systematic errors, a clear sky model aiming to get a more stationary time series were used.

Using clear sky models, the relative error of global forecasting improved by nearly 5%. The results showed that the model is more accurate in clear sky days. While rRMSE was lower than 30% in July, it is 49% in April because there were more cloudy days.

On the other hand, results of the global power irradiance are better than those obtained for titled irradiance. As the authors commented, "the forecast errors are amplified by conversion of the irradiance on the tilted plane".

In fact, values were slightly different. For global power irradiance, rRMSE was 28% in July and 44% in April versus the values referred above of 30% and 49%, respectively.

A three step model was described in *Lorenz and Scheidsteger* [24]. Initially, in order to predict global irradiance, a "spatial averaging procedure is applied" using ECWMF derived forecast and measured irradiation data. Then a physical based clear sky model is combined with the first step resulting in forecast aiming to represent the daily solar course. These two steps led to a systematical overestimation for "intermediate cloud cover". Then, overestimation was corrected by applying a polynomial correction and the final prediction was provided.

A modified up-scaling model was described where geographical location dependence weighted the contribution of the several PV systems leading to the regional power production forecasting. Table 2.4 resumes the RMSE for the control area.

Table 2.4 – Results obtained in [24]

| | RMSE | |
|---|---|---|
| | **Intra-day** | **Day-ahead** |
| **Persistence** | **7.4%** | **9.2%** |
| **Operational** | **5.1%** | **4.9%** |
| **Modified up-scaling** | **4.3%** | **4.3%** |

*Chen et al*. [25] proposed a method based on Radial Basis Function Network (RBFN) combined with self-organizing maps (SOM) for one day-ahead solar power predictions. In this model, day and month, daily power output at moment t, mean daily values obtained from NWP of relative humidity H(t+1) (i.e. 24 hour after t moment), wind speed, solar irradiance and air temperature, were considered as inputs.

RBFN is an ANN using a Radial Basis Function (RBF) in hidden layer nodes. It is analogous to use the sigmoid function.

The model was trained in three steps using different techniques: k-mean clustering algorithm, nearest-neighbor method and least square.

According to the weather conditions, SOM divides dataset in three cases: "sunny, cloudy and rainy".

Regarding the Mean Absolute Percentage Error (MAPE) (defined in Eq. (2.8)), the predictions accuracy varied from 6.36% for a cloudy day to 54.44% for a raining day in January.

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\frac{|\widehat{p_t} - p_t|}{p} \times 100\,\% \tag{2.8}$$

*Sfetsos and Coonick* [26] proposed the use of a ANN model where the mean hourly solar radiation measures, obtained from Corsica, was taken as an input to produce one day-ahead forecast. In contrast to similar studies, data was not pre-processed.

Several univariate and multivariate ANN were trained and the results showed an improvement over the Persistence model. While multivariate models improved in a range from 66.38% to 74.04%, univariate case reached in a range from 64.64% and 71.94%. Moreover, while the ARMA univariate presented an improvement of 62.89%, the linear regression multivariate obtained an improvement of 60.91%.

*Neves* [4] proposed a five-stages process to produce up to 72 hours-ahead forecasts. The first step used analytical models to find the extraterrestrial irradiation as a function of day of the year and hour of the day. ANN used as input NWP and extraterrestrial irradiation calculating both diffuse and horizontal plane irradiation. In order to achieve solar power forecasting used different ANN models, in subsequent steps.

Referring to percentage normalized mean absolute error (NMAE) (described on Eq. (2.9) the author achieves 5.59% for 3 days in December and 6.36% for 3 days in January.

$$NMAE = \frac{\frac{1}{N}\sum_{t=1}^{N}|p_t - \widehat{p_t}|}{\overline{\frac{1}{N}\sum_{t=1}^{N}p_t}}x100\ (\%) \tag{2.9}$$

*Silva* [27] compared the performance of ANN, Support Vector Machines (SVM) and ARIMA models up to 72 hours-ahead. While the first two models take NWP as an input, ARIMA model only uses past values of the time series.

The improvement of these models was calculated with respect to the Persistence model (based in Eq. (2.10). Best results were obtained on winter season (from November to January), where the ANN achieved an improvement of 23.73%. In summer (from July 21st to September 21st) and maybe supported by the higher weather stability, AR was the right model for fitting with and improvement of 45.55%.

$$Imp_{NMAPE_k}^{ref} = \frac{NMAPE_k^{ref} - NMAPE_k}{NMAPE_k^{ref}} \tag{2.10}$$

### 2.3.3 Spatial-temporal Forecast

In order to provide clouds behavior information, spatial-temporal forecast have been suggested as an alternative to satellite-based models. The use of historical data from neighboring, in addition to historical data of the forecasting site, characterizes these models.

In contrast to the model described in sub- section **Error! Reference source not found.**, hich is univariate (only uses past values from the same time-series); there are models that combine information from neighboring sites. *Berdugo et al.* [28] benefited from neighboring information by suggesting a kNN model. The authors compared analog collaborative versus

linear regression both collaborative and non-collaborative. One of their goals was to keep data privacy while a data mining process was acquired. Then, it was showed that in the kNN method, the PV site generating data is the only option with access to local power production history.

Authors considered that if the present moment is almost equal to other one of the past, then the near future should be similar to what happened in past. Therefore, to predict future values used the identification of similar situations in the past.

Based on this approximation, each site, in its historical data, searched an analogous situation to the present. As soon as the site identifies an analog moment in the past, the site sends the information to neighboring sites. If predictions from neighboring sites are similar, these predictions would be considered stronger than those which are not similar.

The presence of a similar global situation in the past indicates the strength of area forecast.

The error difference between collaborative and non-collaborative methods indicated that collaboration can be effectively improved the forecast.

Although linear models, especially collaborative one, presented better results than kNN collaborative models, it was not able to solve restriction problems such as the amount of data which should be transmitted or kept confidential since this model requires centralized computation.

*Yang et Xie* [29] proposed an ARX method that makes use of information from 3 different sites and forecasts for 5 minutes-ahead (5 min), 15 mininutes-ahead (15 min), one hour-ahead were produced. The data related with one hour-ahead forecasting were divided in 4 time periods (January 1st/March 31st, April 1st/June 30th, July 1st/September 30th and September 1st/November 30th).

Table 2.5 - Performance evaluation for one hour-ahead forecast[29]

| | RMSE% | |
|---|---|---|
| **Index** | **ARX** | **Persistence** |
| **Case 1** | **12.08** | **14.89** |
| **Case 2** | **15.54** | **18.92** |
| **Case 3** | **16.49** | **29.21** |
| **Case 4** | **15.81** | **28.30** |

Table 2.5 shows that the Spatial-temporal model presents better results than the Persistence model for all cases. However, the improvement of ARX over Persistence does not show a clear tendency since the improvement in autumn and summer months is better than in winter and spring seasons.

In day-ahead forecast, the improvement of spatial-temporal over PSS varied between 3.98 and 6.91.

$$improvement = Persistent_{RMSE} - ARX_{RMSE} \quad (2.11)$$

## 2.4 Final remarks

The behavior of weather especially cloud motion is critical in solar power forecast. In order to introduce information over cloud motion, both NWP and Satellite images are been used in models, as described in sections 2.1 and 2.2.

For very-short time horizons, models that only include historical data showed to be as good as NWP based models and Satellite image, based models. However, the inclusion of NWP or Satellite-image data improved the performance of historical data based models on short term horizons.

Aiming introduce cloud motion information, spatial-temporal methods, such as those described in section 2.3.3, use information of neighboring solar sites. The analysis conducted in this section shows that only two studies were developed with spatial-temporal techniques referring to distributed PV power forecasting. One study uses a kNN based method, but it does not focus in getting the best forecast. The goal is to maintain data privacy and handle data streaming of solar power time series. On the other hand, the second method proposed an AR model but the authors did not use a VAR model like it is described in this dissertation.

Statistical models based on past values have shown better results than NWP and Satellite image based models for very short term forecasting. In this dissertation, we aim to improve the results by using VAR model and VAR with exogenous inputs (VARX) because both alternatives allow us to include past data from neighboring sites in order to emulate information about cloud motion.

Then, achieving the best forecast using information from nearby sites constructed on the top of smart grid architecture is the new contribution of this dissertation.

# Chapter 3

# Spatial-temporal forecasting methodology

This Chapter describes the models developed in this thesis, related to very short term forecasting of solar photovoltaic power injection, in a grid with the structure and organization of a smart grid including dispersed generation. It is understood that very short-term forecasting is very important for the Distribution System Operator (DSO) and energy market agents; moreover, models of the kind of vector autoregressive (VAR), which will be the core of the thesis work, require central computation – but the raw data is collected from the distributed sources connected to the grid.

The *"Évora Inovcity"* project is an innovative project from *"EDP Distribuição"* in collaboration with *"EDP Inovação"*, *"INESC Porto"*, *"EFACEC"*, *"LOGICA"* and *"JANZ/CONTAR"*. This project aims to contribute to a more efficient grid operation, improving the Quality of Service (QoS), decreasing costs, leading to a rise of environmental sustainability.

The growing of liberalized electricity markets emphasizes the advantages of a more flexible and efficient grid operation, evidenced by the emergence of tariffs that are more flexible. The evolution of electricity markets and the smart grid concept should lead to the emergence of new services and products adapted to the consumer. Note that with micro generation growing the consumer evolves to a producer/consumer paradigm.

In *"Évora Inovcity",* the Energy Box (EB) replaces the classic energy meter, allowing the consumer access to real-time and detailed information about his consumer habits. For instance, client can know the hourly energy consumption. Due to the financial impact, the access to this real-time information can lead a change to more efficient consumption habits.

EB and the communication processes play a key role in "Évora Inovcity" Smart Grid. Section 3.1 presents a description of the context surrounding the forecasting problem and the communication processes in place.

In forecasting problems, the goal is to estimate values (also designated as output values or response values), generally designated by $y$ in the forecasting literature, but here referred to as $p$, where $p$ are photovoltaic production (PV) values.

In order to solve the prediction problem, normally a set of random output variables and a set of input or explanatory variables, generally designated by $x$, but here referred to as $p_{(t-l)}$, constitute a training set, where $p_{t-l}$ are past PV values.

The training set is used to develop a model mapping $p_{(t-l)}$ to $p$.

Traditional models like Autoregressive (AR) and Vector Autoregressive (VAR) assume that the data series present stationarity. The Solar data series present seasonality both in hour of day and day of the year. Section 3.2 describes a clear sky model to provide stationarity to data series.

The following, presents the AR models and the VAR models, as well as their application to the solar power distributed forecasting. The Spatial-temporal models section is divided into two subsections.

Section 3.4.1 presents the VAR models, which only use the same type of input from other similar data series. While, Section 3.4.2 explains the philosophy and the operationalization of VAR with exogenous variables (VARX) models, including the inputs used in VAR models and those from data series of distinct types.

Consecutively, section 3.5 describes the online Recursive Least Squares fit (RLS) approach.

Section 3.6 explains the Gradient Boosting (GB) method, in both univariate and multivariate versions, and the Cross Validation method (CV). In order to estimate the optimal iterations number to use in the GB algorithm, aiming to reduce the error, avoiding overtraining, we adopt the cross validation (CV) algorithm.

## 3.1 Communication Infrastructure of the Smart Grid

The concept of smart-grid aims to provide more flexibility to the management of the electricity system, namely in the low voltage level. Smart-grids require a communication infrastructure that enables information exchange among the several network components.

Figure 3.1 illustrates an overview of the three-layer communication system present in Évora Inovgrid.

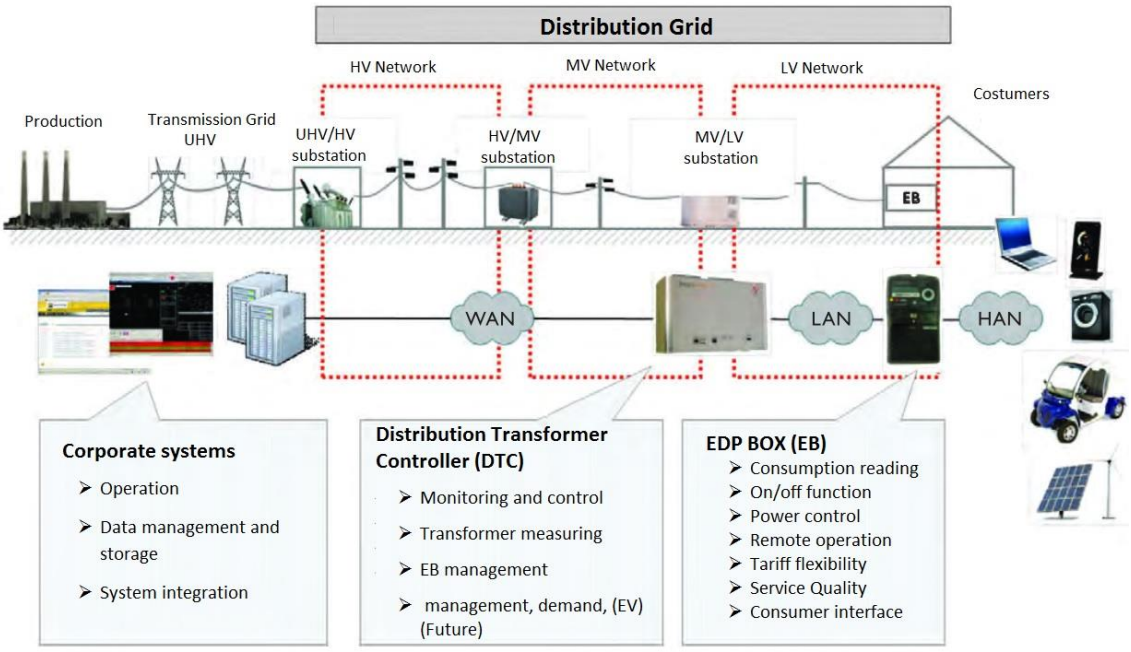


Figure 3.1 Inov-Grid architecture adapted[2]

The major power system innovation provided by a smart grid is expected at the low voltage (LV) level. As mentioned, one key element in the Inovcity project is de EB, *Monteiro* [2]. An EB

is a piece of equipment with several capabilities, among which the possibility of measuring separately both load and generation.

The EB can communicate with client devices using different technologies, such as Wi-Fi, ZigBee or Power Line Carrier (PLC).

The EB are connected to the Distribution Transformer Controller (DTC), which is a device installed at the secondary substation (distribution transformer) level, i.e. on the Medium Voltage/Low Voltage (MV/LV) transition.

Monitoring and control of the secondary substation, information analysis, reporting and acting locally and concentration of information collected from EBs are some of the DTC functions.

The communication between DTC and EB is performed mainly over PLC, PLC Prime or as complementary technology General Packet Radio Service (GPRS). Some tests over Local Area Network (LAN) have been run based on RF Mesh technology.

The communication between the DTC and Power system is based on a Wide Area Network (WAN) and is supported by GPRS/UMTS (Universal Mobile Telecommunications System) connectivity, although wire based technology such as Asymmetric Digital Subscriber Line (ADSL) or optical fiber can be used.

The connection between EB and DTC and the remaining power system is crucial to the operation of the models presented in this dissertation, since the Vectorial Autoregressive (VAR) models require central processing, at Distribution System Operator (DSO) level, for example.

## 3.2 Clear sky models

Auto-regressive (AR) and vector auto-regressive (VAR) models assume data stationarity. Solar power production presents seasonality, on both hour of the day and day of the year.

Statistical models use a clear sky model to attenuate seasonality, as referred to in section 2.3. In the literature, two different clear sky models classes can be found: deterministic or statistical.

Deterministic models are based on physical parameters such as sun elevation angle and sun azimuth angle.

Statistical models use historical values of solar irradiance or solar power generation and statistical smoothing techniques, in order to estimate clear sky value for each hour of each day of the year.

Following the results presented by *Bacher el al*. [16], a statistical clear sky model is mentioned in this section.

Measurements of solar power production, taken over a year, are used as input of a weighted quantile regression function (wqr), to estimate solar power production on a clear sky day ($\hat{p}_t^{cs}$). The other inputs of wqr are: hour of the day ($h$), day of the year ($doy$), size of the kernel for the hour ($sh$), size of the kernel for the day ($sdoy$) and the quantile probability ($\tau$).

The wqr is used to calculate $\hat{p}_t^{cs}$. Assuming a zero order polynomial, the wqr inputs are: solar power production, weights ($W$) and quantile probability ($\tau$).

As mentioned, the clear sky power values are used to remove the data series seasonality. Using Equation (Eq.) (3.1), the normalized solar power production for each hour of each day is calculated, aiming at removing the dataset seasonality.

$$p_t^{norm} = \frac{p_t}{\hat{p}_t^{cs}} \tag{3.1}$$

with $t = f(h, doy)$,

where:

$p_t$ is solar power production measured at each couple of $(h)$, $(doy)$ variables,

$\hat{p}_t^{cs}$ is the solar power production estimated for each couple of $(h)$, $(doy)$ variables on clear sky days, and

$p_t^{norm}$ is the normalized value calculated.

Due to the small solar power production both at sunset and sunrise, $\hat{p}_t^{cs}$ can be a small value. To avoid increasing the errors, a selection criterion must be defined. One could choose to ignore all $p_t^{norm}$ in which $\hat{p}_t^{cs}$ was lower than a reasonable value, e.g. 0.2. This option can lead to a different sample size for the several sites.

Since in this dissertation VAR models are used and these require equal sample size for every sites, the criterion used is that solar power production between 20h and 6h are considered not available (value set to NA).

## 3.3 Auto-regressive models

Auto-regressive integrated moving average (ARIMA) models are forecasting models based on the box-jenkins methodology.

It is common to refer to ARIMA models as ARIMA(l,m,n) were l, m and n are non-negative integers. Parameter l represents the order of auto-regressive (AR) component of model, while m refers to model integrated component (I) order and n is the model moving average (MA) component. An ARIMA(l,0,0) is also known as AR(l).

AR models provide a good starting point for constructing forecasting models based on historical values of solar power generation.

Distinct parameters and models may be used, for each site and time horizon, since the models are adjusted separately for every time horizon and for each data series, i.e. each DTC or EB historical data.

The goal of this dissertation is to compare forecasting methods, up to six hours-ahead.

To achieve this objective, a variant of a classical AR model with l order is applied. In addition to the l order lag variables, the inclusion of a seasonal term is also considered, in the model construction [30].

Equation (3.2) defines an AR model, including two lag variables, and a seasonal term, for one hour-ahead.

The two lag variables considered are: the last known value and the previous hour value, i.e., $(t)$ and $(t-1)$, while, the seasonal term is the PV production recorded 24-hours before the time for which the forecast is being made, i.e., $(t-24+k)$ where $k$ is the forecasting time horizon.

$$p_{t+1} = \alpha + \beta_1 p_t + \beta_2 p_{t-1} + \beta_{(l+1)} p_{t-l} + e_{t+1} \tag{3.2}$$

where:

$\alpha$ is the intercept coefficient,

$\beta$ are the model coefficients and

$e_{t+1}$ is the error, assumed to be white noise with mean zero and fixed variance $\sigma_e^2$

The autocorrelation function (ACF) is used to ascertain the most important lag variables that should be used in model.

An evident exponential decreasing, in first few hours, was found by *Bacher* [31] after plotted ACF of the normalized solar power production against lag variables. This behavior suggest an AR(1) component.

In the same plot, a seasonal lag of 24 term can be identified, due to the same exponential decreasing behavior.

Figure 3.2 and Figure 3.3 show, both for a DTC and an EB, respectively, ACF applied to $p_t^{norm}$ data set, used in chapter 4. Both figures show a clear correlation with the last known hour, $p_t$.

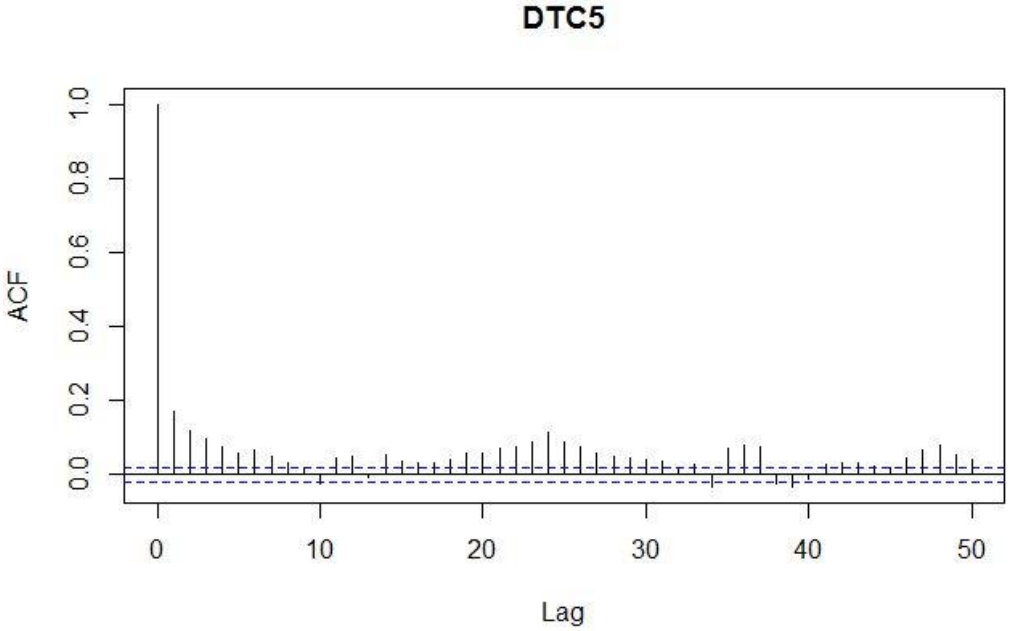


Figure 3.2 – DTC ACF for first 50 lag variables of $\hat{p}_t^{norm}$

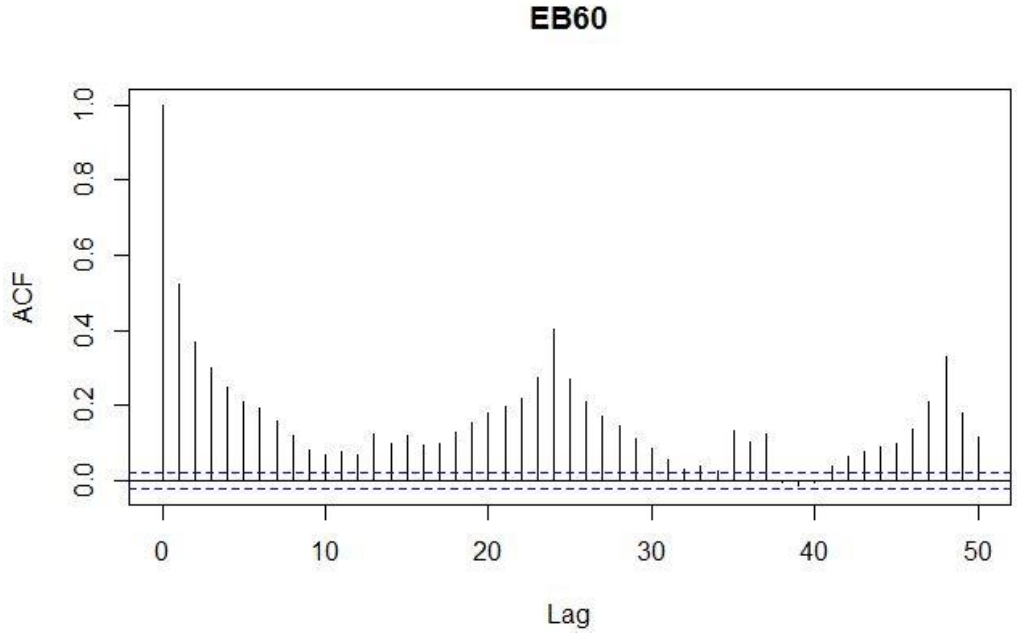


Figure 3.3 – EB ACF for first 50 lag variables of $\hat{p}_t^{norm}$

As clear sky model is not capable of removing all seasonality of data series, i.e. $\hat{p}_t^{norm}$ are not stationary, Figure 3.2 and Figure 3.3 show a peak at season lag of 24 and 48 hours.

Since a peak is in previous hour variable $(t-1)$ , models developed throughout this dissertation use the variable as input.

After Figure 3.2 and Figure 3.3 analysis, it was decided that the AR model to be used would be an AR model with two lag variables and one seasonal term.

Note that while the lagged variables used as input are the same, for the whole time-horizon, the variable seasonal input differs with the time horizon.

The models used, in the training step, from one hour-ahead to six hour-ahead are identified in Eq. (3.3) to Eq. (3.8):

$$\hat{p}_{t+1} = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + \beta_3 p_{t-23} + e_{t+1} \tag{3.3}$$

$$\hat{p}_{t+2} = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + \beta_3 p_{t-22} + e_{t+2} \tag{3.4}$$

$$\hat{p}_{t+3} = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + \beta_3 p_{t-21} + e_{t+3} \tag{3.5}$$

$$\hat{p}_{t+4} = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + \beta_3 p_{t-20} + e_{t+4} \tag{3.6}$$

$$\hat{p}_{t+5} = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + \beta_3 p_{t-19} + e_{t+5} \tag{3.7}$$

$$\hat{p}_{t+6} = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + \beta_3 p_{t-18} + e_{t+6} \tag{3.8}$$

where

$p_{t+k}$ are forecast values for k hours-ahead,

$\beta_0$ are the intercept values, i.e., a constant value,

$\beta_i$ (with $i \neq 0$) are the coefficients for the different time lag variables, i.e. $\beta_1$ are the coefficients for the actual hour value, $\beta_2$ are the coefficients for the hour before value and $\beta_3$ are the coefficients to be applied to the measured value 24h before the forecasting time, and

$e_{t+k}$ are the residuals for each forecasting time horizon.

Since they refer to an AR model, all the Eq. from (3.5) to (3.8) apply individually to each DTC or EB dataset.

Forecasting using AR model is a two steps procedure. Firstly, the model has to be trained, in order to get the $\beta$ coefficients (response coefficients). Aiming that a training data set is normalized, using Eq. (3.1), composed of solar power productions measured between 2011 February 1st and 2012 January 31st.

Then, for each hour of each day, coefficients are applied to normalized solar power productions lag variables.

A test data set, composed by solar power productions measured between 2012 January 31st and 2013 March 6th is normalized using Eq. (3.1).

Finally, as described in section 3.5, using the known values as target applies the least-squares fit.

Considering the linear model presented in Eq. (3.3) as forecast model, it is possible to select the coefficient from which minimizing the error.

Equation (3.9) presents a generalized expression of Eq. (3.3).

$$\hat{P}_{(t+1)} = \mathrm{B}^{\mathrm{T}} P_{t-l} + e \tag{3.9}$$

where:

$\mathrm{B}$ is the response coefficients vector,

$P_{t-l}$ is the vector which includes the lag variables vector.

The response coefficients can be chose based on which minimize the error of Eq. **Error! eference source not found.** and (3.9). Eq. (3.10) shows a criterion function.

$$V_N(\mathrm{B}) = \frac{1}{N}\sum_{1}^{N}(p - \mathrm{B}P_{t-l})^2 \tag{3.10}$$

The $V_N(\mathrm{B})$ criterion has quadratic dependence on $B$ so it can be minimized using Eq.(3.11) analytical expression,

$$\hat{\mathrm{B}}_N = (\,(P_{t-l})(P_{t-l})^T\,)^{-1}(P_{t-l})^T P_{t+1} \tag{3.11}$$

since the inverse exists.

This is the well-known ordinary least square fit (OLS).

The training step by the application of the OLS returns the predictors coefficients. Taking Eq. **Error! Reference source not found.** as example, the training model will provide $\beta_0$, $\beta_1$ ,$\beta_2$ nd $\beta_3$. According Eq. (3.12) these coefficients are applied to test data set.

$$\hat{p}_{t+k} = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + \beta_3 p_{t-23} + e_t \tag{3.12}$$

where:

$\hat{p}_{t+k}$ are the forecasted values, for k hours-ahead,

$\beta_0$ are the intercept values, i.e a constant value,

$\beta_i$ $(i \neq 0)$ are the coefficients for the different time lag variables, i.e. $\beta_1$ are the coefficients for the actual hour value, $\beta_2$ are the coefficients related with the previous hour and $\beta_3$ are the coefficients to be applied to the measured value 24h before the forecasting time, and

$e_{t+k}$ are the residuals for each forecasting time horizon.

Then, the coefficients are applied to the lag variables of the test data set, in order to achieve the forecasting PV production. Thus, the $\hat{p}_t^{norm}$ lag variables are obtained. Next, using the response coefficients, obtained in the training step, according Eq. (3.3) and (3.8), it is possible calculate the forecast normalized values.

Finally, Eq. (3.13) shows how to obtain forecast, in power values, from predicted values.

$$\hat{p}_{t+k} = \hat{p}_t^{cs}\hat{p}_{t+k}^{norm} \tag{3.13}$$

## 3.4 Spatial-temporal models

As Spatial-temporal models, like VAR and VARX, use neighborhood information, is expected an improvement, over AR, in forecasting accuracy.

One weakness of the AR models is that the AR models do not allow use of information from other correlated time series. VAR models come from a generalization of AR models; however, VAR models allow use of more than one variable.

In fact, since the predictors compute separately for each response dataset, a VAR model is like an AR with exogenous variables (ARX) model. This ARX model uses as exogenous variables the same time lag variables used as endogenous, derived from correlated sites.

Unlike what happens with the AR method, VAR method trains the various data sets together. However, it calculates distinct coefficient for each data series.

The VAR models normally include different values in response coefficients for each data series. However all variables are seen as symmetric.

From the literature analysis (section 2.4), it can be concluded that for very short-term forecasting, up to four hour-ahead, past values are the most important information in solar power forecasting.

### 3.4.1 Vector Autoregressive

The knowledge gained from the study of the AR model is a good starting point for the development of the Vector Autoregressive (VAR) model.

Thus, also for the VAR model, initially, the last known solar power production $(t)$ and the previous hour relative to the time at which the prediction is performed $(t-1)$ are lagged variables , while the same hour of the previous day relative to the time for which the forecast is made $(t-24+k)$, is used as seasonal variable.

VAR model forecasts separately for every time horizon. However, for every time horizon, all predictions are made in a single step. Therefore, although the parameters may have different values, the models have to be the same for all forecasting targets [32].

One advantage of the VAR model, as mentioned, is to allow using information from a neighborhood.

In this thesis, VAR models are considered with $n$ observations of a $q$-dimensional response and $p$-dimensional predictor.

Using Eq. (3.14) to determinate $p$ and assuming $z$ lag variables for each predictor and a system with $q$ different datasets, the VAR model calculates a predictor matrix of $pq$ dimension, for each time horizon.

$$p = (zq + 1) \tag{3.14}$$

The mathematical expression used to calculate the predictor coefficients for “DTC1” and “DTC2”, assuming DTC forecasting, for one hour-ahead, are described in Eq. (3.15) and (3.16), while in Eq. (3.17) and (3.18) similar expressions are presented, for “DTC1” and “DTC6”, but for four hours-ahead.

$$\begin{aligned} p_{t+1,1} = \beta_{0,1} + \beta_{1,1}p_{t,1} + \beta_{2,1}p_{t-1,1} + \beta_{3,1}p_{t-23,1} + \beta_{4,1}p_{t,2} + \beta_{5,1}p_{t-1,2} + \cdots \\ + \beta_{28,1}p_{t,10} + \beta_{29,1}p_{t-1,10} + \beta_{30,1}p_{t-23,10} + e_{t+1,1} \end{aligned} \tag{3.15}$$

$$\begin{aligned} p_{t+1,2} = \beta_{0,2} + \beta_{1,2}p_{t,1} + \beta_{2,2}p_{t-1,1} + \beta_{3,2}p_{t-23,1} + \beta_{4,2}p_{t,2} + \beta_{5,2}p_{t-1,2} + \cdots \\ + \beta_{28,2}p_{t,10} + \beta_{29,2}p_{t-1,10} + \beta_{30,2}p_{t-23,10} + e_{t+1,2} \end{aligned} \tag{3.16}$$

$$p_{t+4,1} = \beta_{0,1} + \beta_{1,1}p_{t,1} + \beta_{2,1}p_{t-1,1} + \beta_{3,1}p_{t-20,1} + \beta_{4,1}p_{t,2} + \beta_{5,1}p_{t-1,2} + \cdots + \beta_{28,1}p_{t,10} + \beta_{29,1}p_{t-1,10} + \beta_{30,1}p_{t-20,10} + e_{t+4,1} \quad (3.17)$$

$$p_{t+4,6} = \beta_{0,6} + \beta_{1,6}p_{t,1} + \beta_{2,6}p_{t-1,1} + \beta_{3,6}p_{t-20,1} + \beta_{4,6}p_{t,2} + \beta_{5,6}p_{t-1,2} + \cdots + \beta_{28,6}p_{t,10} + \beta_{29,6}p_{t-1,10} + \beta_{30,6}p_{t-20,10} + e_{t+4,6} \quad (3.18)$$

where:

$p_{t+k,j}$ are the target values, in training step, or forecast values, in test step, for $k$ hours-ahead for predictor $j$,

$\beta_{0,j}$ are the intercept values for predictor $j$, i.e a constant value,

$\beta_{i,j}$, $(i \neq 0)$ are the coefficients for the different time lag variables for each predictor, and

$e_{t+k,j}$ are the residuals for $k$ hours-ahead for predictor $j$.

Note that $\beta_{0,1}$ from Eq. (3.15) and (3.17) although have the same meaning in both equation, $\beta_0$ represents the intercept term of univariate regression, i.e. a constant value, $\beta_0$ assumes different values in each equation.

The same explanation can be used relatively to $\beta_{1,1}$ also present in Eq. (3.15) and (3.17). This coefficient represents, in both equations, the contribution of present solar power production for "DTC1" forecasting, obviously for different time horizons.

The term $p_{t-1,1}$ present in Eq. from (3.15) to (3.18), which represents the solar power production measured one hour before, for DTC 1, take the same value in all four equations.

As mentioned, the VAR model brings a clear influence of the information from neighboring sites, to the forecasting model.

The expression presented in Eq. (3.19) it is a VAR model generalization, in matrix notation.

$$P_{t+1} = BP_{t-l} + E \quad (3.19)$$

where:

$P_{t+1}$ it is a $Nq$ dimension matrix, of target values, in the model development step, or predicted values, in the forecasting step,

$P_{t-l}$ it is a $Np$ dimension matrix, containing a column of 1 and the lag values

$B$ it is a $pq$ dimension matrix of response coefficients, and

$E$ it is a $Nq$ dimension matrix containing the errors, which are assumed to be white noise with mean zero and fixed variance $\sum_e$

Note that the $P_{t-l}$ matrix it is the same for all DTC or for all EB, depending on which are been forecasted, but is different for each time-horizon.

OLS applying the Eq. (3.20) can estimate the parameters.

$$\hat{B} = P_{t+1}P_{t-l}^T(P_{t-l}P_{t-l}^T)^{-1} \quad (3.20)$$

Since the input variables are the same for all models the multivariate OLS similar to what obtained if the univariate OLS estimator it is applied to each equation separately, *Zellner* [33].

Like in an AR model, the VAR model requires data processing. Firstly, a clear sky model is applied in order to obtain $p_t^{norm}$ values, for all predictors. The next step is to construct the $P_{t-l}$ matrix, which includes the lag variables of all DTC or EB. The first two steps are applied to both the training and test datasets.

After the matrix $P_{t-l}$ construction, in the training stage, like in an AR model, applies the least-squares fit.

Finally, using $\mathrm{B}$ matrix (3.19) the forecast matrix $\hat{P}_{t+1}$ is obtained.

### 3.4.2 Vector Autoregressive with exogenous variables

As mentioned, the VAR models allow introducing, in forecasting models, information from correlated variables. On the other hand, historical values are the most important variables for very short-term forecasting models. However, information about cloud motion proved to be also important.

Aiming to provide information about cloud motion to the forecasting model, Vector Autoregressive with exogenous variables (VARX) model include lag variables from EB, in order to improve the DTC forecasted values. VARX models use EB information, to emulate cloud motion behavior.

The VARX model leads to a significant increase of the response coefficients of matrix $B$. In this case study, assuming three lag variables from each DTC, which are $(t-1)$, $(t-2)$ and $(t-24)$, and two lag variables from each EB, which are $(t-1)$, $(t-2)$, the VARX model uses a matrix $B$ composed by 119 response coefficients for each DTC forecasting model, while the VAR model presents 31 response coefficients for each DTC.

From a mathematical point of view, the VARX model is a VAR model so the forecasting process is very similar.

The used notation is also similar: in VARX models considers $n$ observations of a $q$-dimensional response and $p$-dimensional predictor.

Equations (3.21) and (3.22) describe the mathematical expressions used to calculate the VARX model response coefficient, to predictor number 1 and to predictor number 2, for one hour-ahead. While Eq. (3.23) and (3.24) present the similar expression, to predictor number 1 and predictor number 6 for four hours-ahead.

$$\begin{aligned} p_{t+1,1} = \beta_{0,1} &+ \beta_{1,1}p_{t,DTC1} + \beta_{2,1}p_{t-1,DTC1} + \beta_{3,1}p_{t-23,DTC1} + \beta_{4,1}p_{t,DTC2} + \cdots \\ &+ \beta_{28,1}p_{t,DTC10} + \beta_{29,1}p_{t-1,DTC10} + \beta_{30,1}p_{t-23,DTC10} \\ &+ \gamma_{1,1}p_{t,EB1} + \gamma_{2,1}p_{t-1,EB1} + \gamma_{3,1}p_{t,EB2} + \cdots + \gamma_{88,1}p_{t-1,EB44} \\ &+ e_{t+1,1} \end{aligned} \tag{3.21}$$

$$\begin{aligned} p_{t+1,2} = \beta_{0,2} &+ \beta_{1,2}p_{tDTC,1} + \beta_{2,2}p_{t-1,DTC1} + \beta_{3,2}p_{t-23,DTC1} + \beta_{4,2}p_{t,DTC2} + \cdots \\ &+ \beta_{28,2}p_{t,DTC10} + \beta_{29,2}p_{t-1,DTC10} + \beta_{30,2}p_{t-23,DTC10} \\ &+ \gamma_{1,2}p_{tEB,1} + \gamma_{2,2}p_{t-1,EB1} + \gamma_{3,2}p_{t,EB2} + \cdots + \gamma_{88,2}p_{t-1,EB44} \\ &+ e_{t+1,2} \end{aligned} \tag{3.22}$$

$$\begin{aligned} p_{t+4,1} = \beta_{0,1} &+ \beta_{1,1}p_{t,DTC1} + \beta_{2,1}p_{t-1,DTC1} + \beta_{3,1}p_{t-20,DTC1} + \beta_{4,1}p_{t,DTC2} + \cdots \\ &+ \beta_{28,1}p_{t,DTC10} + \beta_{29,1}p_{t-1,DTC10} + \beta_{30,1}p_{t-20,DTC10} \\ &+ \gamma_{1,1}p_{t,EB1} + \gamma_{2,1}p_{t-1,EB1} + \gamma_{3,1}p_{t,EB2} + \cdots + \gamma_{88,1}p_{t-1,EB44} \\ &+ e_{t+4,1} \end{aligned} \tag{3.23}$$

$$\begin{aligned} p_{t+4,6} = \beta_{0,6} &+ \beta_{1,6}p_{t,DTC1} + \beta_{2,6}p_{t-1,DTC1} + \beta_{3,6}p_{t-20,DTC1} + \beta_{4,6}p_{t,DTC2} + \cdots \\ &+ \beta_{28,6}p_{t,DTC10} + \beta_{29,6}p_{t-1,DTC10} + \beta_{30,6}p_{t-20,DTC10} \\ &+ \gamma_{1,6}p_{t,EB1} + \gamma_{2,6}p_{t-1,EB1} + \gamma_{3,6}p_{t,EB2} + \cdots + \gamma_{88,6}p_{t-1,EB44} \\ &+ e_{t+4,6} \end{aligned} \tag{3.24}$$

where:

$p_{t+k,j}$ are the target values, in training step, or forecast values, in test step, for $k$ hours-ahead for predictor $j$,

$\beta_{0,j}$ are the intercept values for predictor $j$, i.e a constant value,

$\beta_{i,j}$, $(i \neq 0)$ are the coefficients for the different DTC time lag variables for each predictor,

$\gamma_{i,j}$, are the coefficients for the different EB time lag variables for each predictor, and

$e_{t+k,j}$ are the residuals for $k$ hours-ahead for predictor $j$.

Note that, while $\beta_{i,j}$ are normalized solar power production values, $\gamma_{i,j}$ are solar power production measurements.

Equation (3.27) presents the VARX model generalized matrix notation. Making

$$\mathrm{B} = \alpha + \beta_1 + \beta_2 + \cdots + \beta_{30} + \gamma_1 + \gamma_2 + \cdots + \gamma_{88} \tag{3.25}$$

and

$$\begin{aligned} P_{t-l} = 1 + p_{t,DTC1} + p_{t-1,DTC1} + p_{t-23,DTC1} + \cdots + p_{t,DTC10} + p_{t-1,DTC10} \\ + p_{t-23,DTC10} + p_{t,EB1} + p_{t-1,EB1} + \cdots + p_{t,EB44} + p_{t-1,EB44} \end{aligned} \tag{3.26}$$

$$P = BP_{t-l} + E \tag{3.27}$$

where:

$P$ it is a $Np$ dimension matrix, of target values, in the model development step, or predicted values, in the forecasting step,

$X$ it is a $Nq$ dimension matrix containing both normalized lag values from the DTC and measured lag values from the EB,

$B$ it is a $pq$ dimension matrix of response coefficients, and

$E$ it is a $Nq$ dimension matrix containing the errors, which are assumed to be white noise with mean zero and fixed variance $\sum_e$

The VARX model requires data processing. Firstly, a clear sky model is applied, in order to obtain $p_t^{norm}$ values, to all predictors. The next step is to construct the $P_{t-l}$ matrix, which includes the DTC normalized lag variables and EB measured lag variables, either normalized or not. The first two steps are applied to both training and test datasets.

In the training stage, after the matrix $P_{t-l}$ construction the ordinary least-squares regression function is applied.

Finally, applying the $B$ matrix to the lag variables of test dataset VARX model provide forecast values.

## 3.5 Recursive least-squares fit

As well known, the efficient operation of the electrical system depends on an accurate knowledge of several variables such as load demand, wind power production, solar power production, among other.

The very short term forecasting is of particular importance to “on-line” decision support, i.e., decisions made during the system operation. Therefore, given that new data is generated during the system operation, the models should be able to adapt to new data.

Recursive algorithms allow update of the models in order to incorporate real-time data in models.

In on-line models, the model structure must be chosen initially; however, there is no guarantee that the new data will not change problem structure. In this, the model chosen could no longer be the best one.

The Recursive Least-squares regression fit (RLS) is described in this section.

Aiming to support real-time decision and incorporate real-time data, in recursive identification it is important that memory space and computation time do not raise over time, *Lennart and Sörderström* [34]. However, this dissertation will not provide this kind of analysis.

The generic recursive identification models objective is to forecast, for each discrete time considered $(t)$, a value $\hat{\beta}(t)$.

In order not to allow that computational time increases with data sample, the data needs to have fixed size auxiliary "memory" quantity. Foregoing memory updates according to the algorithm structure.

*Lennart and Soderstrom* [34] describe in detail the generic recursive identification mathematical model. Mathematical manipulations lead to Eq. (3.28), which is more suitable for computation, since a scalar replaces a matrix inversion.

Note that, as described in sections 3.3 and 3.4, the $q$-dimension (for the base case, i.e., considering as responses two lag terms and one seasonal term, for endogenous variables and two lag terms for exogenous variables) is four for AR models and 31 for VAR models, when DTC are considered. When EB are used, this dimension raises to 133, and for VARX models, the response dimension it is 119.

The presented version includes a "forgetting factor" $(\lambda)$. If any event changes the relation between input variables and output, this $\lambda$ allows, by "forgetting" older information, the model to better adapt to the changes. If $\lambda = 1$ is used then the RLS with *forgetting factor* is equals to RLS.

$$\widehat{\mathrm{B}}(n) = \widehat{\mathrm{B}}(n-1) + \frac{\Omega(n-1)P_{t-l}(n)}{\lambda + P_{t-l}^{T}(n)\Omega(n-1)P_{t-l}(n)}\Big(p(n) - \widehat{\mathrm{B}}^{T}(n-1)P_{t-l}^{T}(n)\Big), \qquad (3.28)$$

with,

$$\Omega(n) = \frac{1}{\lambda}\left(\Omega(n-1) - \frac{\Omega(n-1)P_{t-l}(n)P_{t-l}^{T}(n)\Omega(n-1)}{\lambda + P_{t-l}^{T}(n)\Omega(n-1)P_{t-l}(n)}\right). \qquad (3.29)$$

where,

$\Omega$ is a symmetric covariance matrix.

The $\Omega$ matrix saves the relation between the different input variables and the output, allowing the model adaptive behavior.

According *Lennart and Sördeström* [34] a large constant diagonal matrix as $\Phi$ matrix and a matrix of zeros as $\mathrm{B}$ matrix can be used as initial values of recursive method. In the scripts used in this dissertation, $\Omega$ initially was a diagonal matrix with the diagonal values equal to 1000.

## 3.6 Gradient Boosting

The Gradient Boosting (GB) is a generic technique for solving regression problems, which allows optimization based on an arbitrary loss function.

In this dissertation, GB will use least squares as the loss function, both for univariate and multivariate GB versions.

The Boosting models presented in this section, both univariate and multivariate, are stagewise versions.

The GB models work like an ensemble of models. Every iteration calculates only the coefficient of the predictor that minimizes the error, considering all the others zero in that iteration. The final model is not more than the sum of the coefficients determined in several iterations.

This is known as stagewise approach and is different from the classical stepwise approach, which readjusts previously entered terms when new ones are added.

As mentioned in sections 3.3 and 3.4 the response coefficient matrix $\mathrm{B}$ can reach large dimensions.

Taking as reference two lag terms and one seasonal term, the AR model matrix $\mathrm{B}$ will present $[1 \times 4]$ dimension. When it consider DTC, the VAR model matrix $\mathrm{B}$ will present $[10 \times 31]$ dimension, while when it use EB the matrix dimension will be $[44 \times 133]$ and for VARX model, the number of predictors for each DTC will be 119, which lead to a $[10 \times 119]$ matrix.

The intention of using VAR models is to provide cloud behavior information to forecasting system. It is expected that several variables have limited or no impact, for a given site forecasting; in these cases the parameters referring to those variables should be zero. With a classic VAR model, this will not happen.

Another problem, associated with the VAR method, is the need to transfer a large amount of information, from all points of measurement (DTC or EB), to the central processing location.

Using GB it is hoped that several predictors will become zero. If a predictor is zero in the end of computational process, his information is not necessary to forecast.

Boosting based models are iteratively constructing, using a “weak learner”, as fit, to supervise the training. A weak learner is a classifier, which is slightly correlated with the response.

The stopping criterion of boosting can be the number of iterations. However, it can lead to overtraining, i.e., reduce the errors in the training set while the test error increases, which, must be avoided.

In order to decide the method used to determine the optimal number of iterations, the Akaike Information Criterion (AIC) and Cross Validation function (CV), were considered.

Although AIC is computationally more attractive then CV according *Lutz and Bühlmann* [35] AIC-based stooping tends to overshoot the optimal number of iterations.

Considering the above, in this section, the sub-section 3.6.3, is devoted to explaining the CV model used to determine the optimal number of iterations $(mstop)$, for both the univariate and multivariate models.

Although, forecasting problems are about minimizing a loss function $(L)$, the choice of a parameterized model allows one to treat the problem as a parameter optimization. Parameters $\mathrm{B}$ that minimize objective function $\Phi$ can be determined using Eq. (3.30), *Friedman* [36].

$$\mathrm{B}^* = \arg\min_{\mathrm{B}} \Phi(\mathrm{B}) \tag{3.30}$$

where

$$\Phi(\mathrm{B}) = E_{p,p_{t-l}} L\big(p, \hat{p}(p_{t-l}; \mathrm{B})\big) \tag{3.31}$$

Appling the calculated coefficients to the input variables the forecasting values are determinate.

Equation (3.32) express solution parameters, for $\hat{p}(p_{t-l}; \mathrm{B})$ and $L$, since numerical optimization is requested to solve Eq. (3.30).

$$\mathrm{B}^* = \sum_{m=0}^{M} \mathrm{B}_m \tag{3.32}$$

where,

$\mathrm{B}_0$ is the initial parameters vector, composed of assumed values, and

$\mathrm{B}_m$, with $1 < m \leq M$ are successive increments (steps or boosts).

The optimization method defines the computational procedures for each step.

The classical Gradient descent method is a simple and often used model in minimization problems. This method, firstly, computes the gradient function, obtaining, for every iteration, the steepest-descent direction, which is defined by the negative gradient $(-g_m)$.

The Gradient Boosting differs from the classical steepest descent method, since derivative performs in the functional space, instead the parameters space.

Applying the numerical optimization principles, the solution in functional space becomes that presented in Eq (3.33).

$$\hat{p}^*(p_{t-l}) = \sum_{m=0}^{N} \hat{p}_m(p_{t-l}) \tag{3.33}$$

Where,

$\hat{p}_0$ is the initial guess vector, and

$\hat{p}_m$, with $1 < m \leq M$ are successive increments (steps or boosts).

Applying optimization in function space to steepest-descent, $\hat{p}_m(p_{t-l})$ is determinate.

It is possible to calculate the gradient function assuming that interchange between integration and differentiation it is possible.

Due to the continuous nature of PV production domain, it cannot estimate the error for all domain points, since it is infinite. To overcome this constraint and increase solution robustness, the model imposes smoothing to the outputs.

Note that the GB paradigm, as presented here, is an algorithm capable of dealing with very large dimensional problems. Thus, is expected that the most important explanatory variables become selected in more iterations, and that these variables provide more explanatory power than simply the inputs.

### 3.6.1 Component-wise gradient boosting univariate

In this section, is described, the univariate GB model, using the linear least squares function as the loss function.

GB univariate (GB uni) approach characterizes by the fact that determining predictors, of a given response, is independent from predictors of the other responses. Thus, although, it considers information from different sites, in the forecasting systems, like in VAR model the prediction coefficients of one response dataset (DTC or EB) are completely independent of the predictors of the other response dataset. However, unlike VAR model, some predictors are zero, and can be eliminated from the forecasting model. Attending the previous, it is possible to classify VAR GB univariate as autoregressive exogenous model (ARX). However, in the beginning, all the models for every dataset are similar, thus one can consider GB as VAR. Here GB is referred as VAR.

The GB problem consists in minimizing an error, like in models previously presented.

Using the gradient function detailed in Eq. (3.34), *Bühlmann* [37] described a component-wise linear least squares GB algorithm capable of minimizing the error. A version adapted to the solar power generation forecasting problem is presented here, and referred as GB univariate Algorithm.

*Bühlman* proposed the using of a shrinkage parameter in updating a fit step. The shrinkage parameter, $\nu$, can be interpreted as the step size when the fit is updated.

According to the author, $\nu$ should range in $]0;1]$. Small $\nu$ values can lead to better values, but increasing the number of iterations that are necessary to reach the optimal forecasting model.

Note that $\nu$ it is an empirical parameter. Thus, distinct values are tested aiming to find the best value.

$$\hat{g}_{(P_{t-l},U)}(x) = \hat{\beta}_{\hat{S}} p_{t-l}^{(\hat{s})} \tag{3.34}$$

with,

$$\hat{S} = \arg\min_{1\le j\le p} \sum_{i=1}^{n} \left(U_i - \hat{\beta}_j P_{(t-l)i}^{(j)}\right)^2 \tag{3.35}$$

and,

$$\hat{\beta}_j = \frac{\sum_{i=1}^{N} U_i {P_{(t-l)}}_i^{(j)}}{\sum_{i=1}^{N} ({P_{(t-l)}}_i^{(j)})^2} \tag{3.36}$$

where

$U_i$ are the residuals,

$\hat{\beta}_j$ is the estimated value for the $j$th predictor in each iteration.

The Least Squares approach presents superior computational properties; therefore, it is a good choice as a fitting criterion, although, other fitting criteria can be used.

As mentioned, using the response values $p$ in boosting algorithms, the loss function is related to the residuals $(u)$.

Taking advantage on linear model properties, each of the $M$ iterations, builts a simple forecasting model. The final forecasting model arises from the junction of several simpler models.

GB univariate Algorithm

1. $F^{(0)} = \bar{p}$

2. $U^{(0)} = P - F^{(0)}$
3. $\mathrm{B} = 0$
4. $For\ m = 1\ to\ M\ do$:
   - 4.1. $\hat{\beta}_j = \frac{\sum_{i=1}^{N} U_i P_{(t-l)_i}^{(j)}}{\sqrt{\sum_{i=1}^{N} (P_{(t-l)_i}^{(j)})^2}}$
   - 4.2. $s = \arg\max_j \hat{\beta}_j$
   - 4.3. $\hat{\beta}_s = \frac{\sum_{i=1}^{N} U_i P_{(t-l)_i}^{(s)}}{\sum_{i=1}^{N} (P_{(t-l)_i}^{(s)})^2}$
   - 4.4. $F^{(m)} = F^{(m-1)} + \nu \beta_s P_{(t-l)_s}$
   - 4.5. $\beta_s = \beta_s + \nu \hat{\beta}_s$
   - 4.6. $U^{(m)} = P - F^{(m)}$
5. $End\ For$

End Algorithm

where,

$F^{(0)}$ is the initial fit vector, which is considered the mean value of measured solar power generation, for the time horizon in forecasting, i.e., from $p_{t+1}$ to $p_{t+6}$,

$U^{(0)}$ is the initial residuals vector, where $P$ is the measures vector, for the time horizon in forecasting,

$\mathrm{B}$ is the predictor coefficients vector, initially a zero vector,

$\beta_j$ is the $j^{\text{th}}$ predictor coefficients vector element, calculated in each iteration,

$\nu$ is the shrinkage coefficient,

$s$ is the predictor coefficients vector element index, which minimize the error in the iteration considered, and

$P_{(t-l)_s}$ is the lag variable or the intercept value vector of the $s^{th}$ predictor.

The GB univariate is valid for random input dataset presenting mean-zero error, thus initially GB univariate function center, at zero, the input variables.

In order to initialize the boosting process, are considered as initial values fit, $F^{(0)}$, residuals, $U^{(0)}$, and predictor coefficients, $\mathrm{B}$. While initial fit and coefficients values are assumed, residuals are a consequence of fit values.

Here, GB univariate uses as initial fit the average response value, and zero as the initial coefficient for all predictors.

Taking as example a GB univariate VAR model composed by 2 response data set, for instance 2 EB, and considering 2 lag variables for each EB, the initial predictors matrix is like show Table 2.1. Note that, as mentioned, in GB univariate method, each Table 2.1 row is updated individually. Thus, in this example GB univariate function runs 2 times, one for each EB.

Table 3.1- Predictors matrix $B$, using 2 EB and to lag input variable per EB

| | $intercept$ | $p_{t,EB1}$ | $p_{t,EB2}$ | $p_{t-1,EB1}$ | $p_{t-1,EB2}$ |
|---|---|---|---|---|---|
| $EB1$ | 0 | 0 | 0 | 0 | 0 |
| $EB2$ | 0 | 0 | 0 | 0 | 0 |

GB univariate Algorithm step 4.1 to 4.3 identifies the coefficient, which minimize the quadratic error and calculate his magnitude.

After this initialization starts the "*for cycle*". For each iteration, the cycle firstly calculates using GB univariate Algorithm step 4.1, the capability of each predictor to update and decrease the forecasting error.

Table 3.2 show the results obtained applying GB univariate Algorithm step 4.1 to Table 2.1 example, in first iteration.

Table 3.2– EB1 predictors applying GB univariate Algorithm step 4.1

| | $intercept$ | $p_{t,EB1}$ | $p_{t,EB2}$ | $p_{t-1,EB1}$ | $p_{t-1,EB2}$ |
|---|---|---|---|---|---|
| $EB1$ | $7.4E-16$ | 1.41 | 1.28 | 1.0 | 1.0 |

Attending the results, the function selects, $p_{t,EB1}$ as, the predictor that reduces more the error, in the first iteration. Note that as the example is for one hour-ahead this is the expected result.

Applying GB univariate Algorithm step 4.3 to $p_{t,EB1}$ predictor and a $\nu = 0.1$ the EB1 predictor vector after one iteration is presented in Table 3.3

Table 3.3– EB1 predictors vector after one iteration for Table 2.1 example

| | $intercept$ | $p_{t,EB1}$ | $p_{t,EB2}$ | $p_{t-1,EB1}$ | $p_{t-1,EB2}$ |
|---|---|---|---|---|---|
| $EB1$ | 0 | 0.076 | 0 | 0 | 0 |

Note that, instead of the residuals, U, one can use the PV generation values, $P$, to select and calculate the magnitude of the coefficient which reduces the most error. However, U, according to literature, reduces the most error.

The linear least squares base procedure determines not only the predictor that minimizes the error but also its magnitude.

The final model is composed of the several simpler models, so the algorithm embeds a coefficient update step. Alternatively, one could record all coefficients calculated in each iteration, and in the final stage compose a model with all coefficients. The update step allows the reduction of the information recorded.

Finally, the model updates residuals, in all iterations.

In the end of the $for$ cycle, one achieves the forecasting values applying the forecasting model to the test dataset.

If a coefficient is never the one that minimizes the error among all the coefficients in any of all m iterations, this coefficient will be zero in the end of boosting process. It is from that characteristic that some sparsity is expected in the predictor coefficient matrix in the end of the GB univariate model.

After applying the GB univariate Algorithm to several EB predictor vectors and joining all the predictor vectors, the predictors matrix is obtained.

### 3.6.2 Component-wise gradient boosting multivariate

The following explains the algorithm of the model developed, here referred as GB multivariate (GB multi) Algorithm.

Gradient Boosting univariate model calculates separately the predictor coefficients for each response data set, such as in OLS based models like the VAR off-line methods, presented in section 3.4.

*Lutz and Bühlmann* [35] theorized a methodology to deal with high-dimensional linear regression problems based on Boosting. Authors proposed a similar method to the one presented previously, in section 3.6.1. However, instead of calculating separately the predictors for each set of responses, the multivariate model determines the predictors simultaneously, for all set of responses. Thus, if in the univariate GB method and in each iteration, is considered a $p-dimensional$, from which only one coefficient is selected, in the multivariate GB model a $qp-dimensional$ matrix is analyzed in each iteration.

Like, referred to in section 3.6.1 GB univariate Algorithm, in GB multivariate Algorithm is tested the shrinkage parameter, referred to in section 3.6.1, aiming to improve the forecasting model.

GB multivariate algorithm is similar to GB univariate; however, presents same differences.

In each Boosting iteration, only the coefficient of one predictor of one response set is modified, and only the fit of the response set affected by the selected predictor is updated. This seems to be a good method in correlated response dataset problems.

The multivariate GB version, unlike what happened in the univariate method, needs a matrix, which relates the several response datasets. This matrix can be the covariance matrix, or in alternative, the identity matrix.

Here referred as gamma $(\Gamma)$ the matrix used to relate the several responses. If one uses identity matrix as $\Gamma$, then Multivariate and Univariate models should achieve very similar results.

It is expected that the sparsity of predictor coefficient matrix increases, using GB multivariate version, relatively to the univariate model, if applied the same number of iterations.

As only one response fit vector is updated in each iteration, this method may need more iterations to reach the optimum model. However, as it is applied just once to achieve the forecasting model of all response data set, considering all the multivariate processing is not necessarily computationally slower than the univariate version.

Like the GB univariate, GB multivariate is valid for random input dataset presenting mean-zero error, thus initially GB multivariate function center at zero the input variables.

The GB multivariate Algorithm is very similar to the GB univariate Algorithm.

The steps are the same, however, while in GB univariate model fit, response and coefficients structures are vectors, in multivariate model this variables are matrix .However, the $P_{t-l}$ is the same in both cases if the same problem is considered.

Like in GB univariate, GB multivariate uses residuals $(U)$, instead PV generation values $(P)$, in steps 4.1 to 4.3 because $U$ reduces more of the error.

The first three steps of GB multivariate function are similar to the GB univariate.

The initialization uses the same principles. Initial fit is the mean value of each response data set, however instead of a vector of N elements, in GB multivariate fit is a $Nq$ dimension matrix, with distinct values in each column.

In GB univariate method, the GB univariate function is called $q$ times, while in GB multivariate method, the GB multivariate function is called just once.

GB multivariate function step 4.1, analogous to what happens in GB univariate function step 4.1, selects the predictor that reduces more of the error in a given iteration. However, in GB multivariate function there are at least the $q$ times more predictors than in GB univariate.

GB multivariate function step 4.1 determines the error reduction for each predictor; GB multivariate function step 4.2 selects that predictor which reduces the error most; and GB multivariate function step 4.3 calculates the magnitude of the coefficient that leads to the maximum error reduction.

Following, in GB multivariate function step 4.4, theoretically, the fit of all responses is updated; however, as only one coefficient in a given iteration is non-zero, just the fit of one response is changed.

Take the same example of the previous section, two response data set, considering two lag variables for each EB the initial. Table 3.4 shows predictors matrix.

If a the coefficient selected is in the first row, only the EB1 changes, while if the coefficient selected belongs to second row, then only EB2 fit changes.

Note that, although, GB multivariate function step 4.4 is written in vector form, it could be written in matrix form, since the coefficient matrix, as mentioned, just presents one non-zero coefficient in each iteration.

Table 3.4- Initial predictors matrix $B$, using 2 EB and to lag input variable per EB

| | $intercept$ | $p_{t,EB1}$ | $p_{t,EB2}$ | $p_{t-1,EB1}$ | $p_{t-1,EB2}$ |
|---|---|---|---|---|---|
| $EB1$ | 0 | 0 | 0 | 0 | 0 |
| $EB2$ | 0 | 0 | 0 | 0 | 0 |

GB multivariate Algorithm

1. $F^{(0)} = \bar{P}$
2. $U^{(0)} = P - F^{(0)}$
3. $B = 0$
4. $For\ m = 1\ to\ M\ do$:
   - 4.1. $\hat{\beta}_{jk} = \frac{\left(\sum_{v=1}^{q} U_v^T P_{(t-l)_j} \Gamma_{vk}^{-1}\right)^2}{P_{(t-l)_j}^T P_{(t-l)_j} \Gamma_{kk}^{-1}},\ (uv \neq jk)$
   - 4.2. $(st) = \arg\max_j \hat{\beta}_{jk},\ (st \neq jk)$
   - 4.3. $\hat{\beta}_{st} = \frac{\sum_{v=1}^{q} U_v^T P_{(t-l)_s} \Gamma_{vt}^{-1}}{P_{(t-l)_s}^T P_{(t-l)_s} \Gamma_{tt}^{-1}}$
   - 4.4. $F_{t,s}^{(m)} = F_{t,s}^{(m-1)} + \nu \beta_{st} p_{(t-l)_s}$
   - 4.5. $\beta_{st} = \beta_{st} + \nu \hat{\beta}_{st}$
   - 4.6. $U^{(m)} = P - F^{(m)}$
5. $End\ For$

End Algorithm

where,

$B$ is the predictor coefficients matrix, with $[qp]$ dimension, initially a matrix of zeros,

$j$, $s$ and $u$ are index referring to the $p$ dimension,

$k$, $t$ and $v$ are index referring to the $q$ dimension,

$F^{(0)}$ is the initial fit matrix, which constructed considering the mean value of each dataset of measured solar power production as the value of each $q$,

$U^{(0)}$ is the initial residuals matrix, where $P$ is the measures matrix for the time horizon in forecasting,

$\beta_{jk}$ is the predictor coefficients matrix element, with $jk$ index, calculated in each iteration,

$s, t$ are the predictor coefficients matrix element index which minimize the error in the iteration considered, and

$\Gamma$ is a matrix which correlate the several responses, in fact can be the inverse of the covariance matrix or simply the identity matrix with $qq$ dimension.

### 3.6.3 Cross Validation

The GB models are iterative modules in which the stopping criterion can be the number of iterations. CV is a method which allows determining the number of GB iterations that minimize, in this dissertation, the quadratic error.

As it is possible to calculate the error in the final of each GB iteration, one could stop the GB iteration when the error increases, relatively to the last iteration.

This approach has two problems: on one hand the method may be trapped at a local minimum; on the other hand, the model could be overtrained.

Since the model, is obtained fitting a training data set, but is used in a different data set (the test data set), in order to get forecasting values, the goal is not getting the minimum error in the training set but in all population.

If one tries to achieve the minimum error with no other constraint it is likely that overfitting occurs, increasing error when the model obtained is applied to a set of different data.

Thus, in additive expansions models, like GB, it is important to find the number of iterations that reduce the population error.

An excessive number of iterations in boosting models can lead to overtraining, and a scarce number of iterations might not produce the best possible forecasting.

In order to avoid this behavior, a cross-validation (CV) model can be used. CV consists in an algorithm that divides the training set into smaller groups. Then, complementary subsets are constructed based on the principle that one subset is used for testing and all the others for training.

The main idea is to use the bigger subset to train the model, and then the smaller subset to represent the population, i.e., if overfitting occurs the error in the small test data set increases and the overfitting is identified.

Running the new subsets in multiple rounds of GB, and using an average over the rounds, reduces variability.

The cross validation developed within this dissertation, applied to the training dataset, is explained in the Cross Validation Algorithm presented below.

The Cross Validation process may use the GB function, however this is not computationally efficient. Thus, during this dissertation, specific versions, on both GB univariate and GB multivariate models, were developed to use in the CV process.

One criterion to stop the cross validation process can be the increase in the error metric used. However, to avoid local minima the strategy adopted is to run for a high fixed number of iterations and select the number of iterations which presents the minimum error.

To the GB univariate, the optimal number of iterations (mstop) is determined separately for each response dataset, but the CV algorithm is the same for both univariate GB and multivariate GB models.

Initially, the original training dataset is split, in equal size groups, according to the defined number of groups $(ng)$. It is important that the size of all groups is the same, or at least very similar, in order to all groups present the same statistic weight.

At this point, the method will produce $ng$ simpler distinguishable models in each of the $m$ iterations. The accuracy of the several "partial" models is evaluated using the test subset.

For each $m$ iteration, the average error, referred to the several response test subsets, is determined and recorded.

Finally, the number of iterations that should be used both in univariate GB and multivariate GB is the one that presents the smaller average error.

As mentioned, while for multivariate GB versions all the models are determined together, in a univariate GB method the model for each response dataset is calculated singly, which allows using different iteration numbers for each response dataset.

Cross Validation Algorithm

1. Split dataset into sub datasets
2. Create new train and tests dataset
3. $error = \infty$
4. $For\ m = 1\ to\ maximum\ number\ of\ iterations$
   - 4.1. $error_{iter} = 0$
   - 4.2. $For\ n = 1\ to\ number\ of\ groups\ (ng)$
     - 4.2.1. Run 1 iteration of GB for train subset n
     - 4.2.2. Save fit, residuals and coefficients of GB train subset
     - 4.2.3. Calculate the $subset_{error}$ for test subset n
     - 4.2.4. $error_{iter} = error_{iter} + subset_{error}$
   - 4.3. $end\ For$
   - 4.4. $if\ (error_{iter} < error)$
     - 4.4.1. $error = error_{iter}$
     - 4.4.2. $iter = m$
   - 4.5. $end\ if$
5. $end\ For$
6. Return $iter$

End Algorithm

# Chapter 4

# Results on case study

This chapter devotes to the presentation and explanation of the results achieved with the models described in previous chapter.

The accuracy of several models is compared when the same data set is used for all models. Section **Error! Reference source not found.** provides a brief description of the tests erformed. Following section 4.1 describes the dataset which have been used for testing the models, including error metrics. Finally, the results analysis is divided in off-line methods, presented in section 4.3, and on-line methods, presented in section 4.4.

## 4.1 Test Case Description

The data used to train and test the models developed throughout this dissertation are composed by 44, of an original sample of 60, one hour time step EB distributed by a 1500 km$^2$ area, and they are aggregated in 10 DTC.

All data series present 18072 observations relating to the hourly PV production between February 1st, 2011 and March 6th, 2013, using UTC.

Considering that all measures between 20 p.m. and 6 a.m. are not available and the models consider $p_t$ and $p_{t-1}$ as input variables, predictions are just available between 9 a.m. and 19 p.m., for one hour-ahead. For two hours-ahead, models can only provide forecasting ranging from 10 a.m. to 19 p.m., while for six hours-ahead, models forecast between 14 p.m. and 19 p.m..

Using the following equation (4.1), the improvement (*imp*) is evaluated in forecasting terms.

$$improvement_k = \frac{RMSE_{k_{ref}} - RMSE_k}{RMSE_{k_{ref}}} 100\ (\%) \tag{4.1}$$

where:

$p_{t+k}$ is the measured solar power k-hours-ahead from prediction time

$\hat{p}_{t+k}$ is the forecasted solar power k-hours-ahead from prediction time

*N* is the prediction sample size.

In order to obtain an individual analysis of the model, the normalized root mean square, nRMSE, was used. Equation (4.2) calculated nRMSE.

$$nRMSE_k = \frac{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(p_{t+k} - \hat{p}_{t+k})^2}}{\max(p_{t+k})} 100(\%) \tag{4.2}$$

where:
$p_{t+k}$ is the measured solar power k-hours-ahead from prediction time
$\hat{p}_{t+k}$ is the forecasted solar power k-hours-ahead from prediction time
*N* is the prediction sample size, and
$\max(p_{t+k})$ is the maximum value measured of PV production at k-hours-ahead from prediction time.

Understanding of the $Bias$ model behavior requires the use of normalized $Bias$ (nBias) as metric and it is calculated according Eq.(4.3).

$$nBias = \frac{\frac{1}{N}\sum_{t=1}^{N}(p_{t+k} - \hat{p}_{t+k})}{\max(p_{t+k})} 100\ (\%) \tag{4.3}$$

where:
$p_{t+k}$ is the measured solar power k-hours-ahead from prediction time
$\hat{p}_{t+k}$ is the forecasted solar power k-hours-ahead from prediction time
*N* is the prediction sample size, and
$\max(p_{t+k})$ is the peak value measured of PV production at k-hours-ahead from prediction time

As in the benchmark models, in this dissertation for both off-line and on-line, AR models were considered. AR models use past measurements of one time series as inputs such as the last known PV production $(p_t)$, the hour before $(p_{t-1})$, and the same hour of day before information, $(p_{t-24+k})$. Throughout this section, this model is referred as AR(2) diurnal, since it is a second order AR with an extra diurnal seasonal term.

In this dissertation, we use several versions of Persistence models allowing us to obtain a naïve reference.

A naïve model is a simple case which normally does not require of computational skills. Naïve model provides a benchmark for most of the more sophisticated methods. Stationary data series often use the Persistence model as a naïve model.

The original Persistence model, referred here as Persistence, is a model that assumes the last known value as the predicted value for any time-horizon (up to six hours-ahead in this dissertation). Equation (4.4) formalizes the Persistence.

$$\hat{p}_{t+k} = p_t \tag{4.4}$$

where
$\hat{p}_{t+k}$ is the forecast value for k hours-ahead, and
$p_t$ is the last known value

A slightly different version of the Persistence model takes as forecast the same PV production that occurred 24 hours before the time for which the prediction is made. Equation (4.5) explains this model, which is referred as Persistence(24).

$$\hat{p}_{t+k} = p_{t-24+k} \tag{4.5}$$

where

$\hat{p}_{t+k}$ is the forecast value for k hours-ahead, and

$p_{t-24+k}$ is the PV value recorded 24 hours before.

Moreover, this text introduces two extra versions of the Persistence model. The main concept is similar to the model Persistence(24). However, these two new versions take as forecasting a value resulting from a two-step process instead of assuming as prediction the recorded value.

According with Eq. (4.6) and as it was explained, in section 3.2, the recorded values are normalized using clear sky function.

$$p_t^{norm} = \frac{p_t}{\hat{p}_t^{cs}} \tag{4.6}$$

with $t = f(h, doy)$,

where:

$p_t$ is the solar power production measured at each couple of $(h)$, $(doy)$ variables,

$\hat{p}_t^{cs}$ is the solar power production estimated for each couple of $(h)$, $(doy)$ variables on clear sky days, and

$p_t^{norm}$ is the normalized value calculated.

Then, according to Eq. (4.7) and Eq. (4.8), it is taken $p_t^{norm}$ or $p_{t-24+k}^{norm}$ as the normalized forecasting values.

$$\hat{p}_{t+k}^{norm} = p_t^{norm} \tag{4.7}$$

where

$\hat{p}_{t+k}^{norm}$ is the normalized forecast value for k hours-ahead, and

$p_t^{norm}$ is the normalized last known value

$$\hat{p}_{t+k}^{norm} = p_{t-24+k}^{norm} \tag{4.8}$$

where

$\hat{p}_{t+k}^{norm}$is the normalized forecast value for k hours-ahead, and

$p_{t-24+k}^{norm}$ is the PV value recorded 24 hours before.

After this step, we use Eq. (4.9), where $\hat{p}_t^{cs}$ combines with the normalized forecasting values, and we determine the forecast value.

$$\hat{p}_{t+k} = \hat{p}_{t+k}^{norm} \hat{p}_{t+k}^{cs} \tag{4.9}$$

Throughout this text, these two versions of Persistence model are designated as Persistence$(\tau)$ and Persistence$(\tau 24)$, respectively.

Figure 4.1 shows a comparison of the different versions of Persistence model at DTC level. The conclusions at EB level are the same, more details can be found in Appendix.

Figure 4.1 shows that one should use $p_t$, i.e., the classic persistence value as reference for one hour-ahead forecasting models. However, for the other time-horizons, Figure 4.1 suggests the use, as reference, of persistence$(24)$, i.e., the PV production of the previous day.

Given that the error for one hour-ahead, in Figure 4.1 is lower for Persistence models using last known value$(p_t)$, we conclude that $p_t$ is the variable which provides more information.

However, as the time horizon increases, the information about what happened in the day before becomes more relevant. The clear sky model has little influence on this variable because the difference of clear sky irradiation for two consecutive days is insignificant.

Note that, although the plot lines of both Persistence$(\tau)$ and Persistence$(\tau 24)$ are overlapping, the $nRMSE$ values obtained are slightly different.

We should remark that, in this text, we exclusively compared Persistence with on-line AR models.

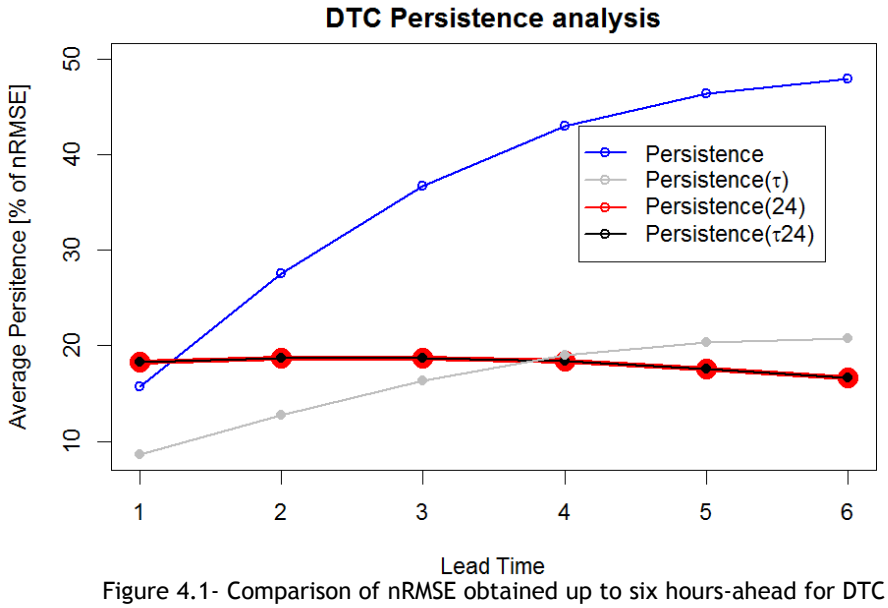


Figure 4.1- Comparison of nRMSE obtained up to six hours-ahead for DTC

## 4.2 Tests

Tested models are divided in off-line and on-line. On-line models, unlike models off-line, are adaptive models, i.e., the coefficients of the models change, as soon as new data becomes available. Section 4.3 presents the offline models analysis, divided in four sub-sections. Initially, section 4.3.1 presents the OLS results. Section 4.3.2 provides an analysis over Gradient Boosting univariate (GB uni) results. In section 4.3.3, presents the Gradient Boosting multivariate (GB multi) results. Finally, section 4.3.4 compares the best results achieved by each model. On the other hand, section 4.4 describes the online models analysis.

Throughout these sections, the AR and VAR models, those take as input variables $p_t$, $p_{t-1}$ and $p_{t-24+k}$, are referred as AR(2) diurnal model, or simply AR, and VAR(2) diurnal model or VAR.. VARX AR(2) diurnal models, or simply VARX, are models which use $p_t$, $p_{t-1}$ and $p_{t-24+k}$ as input variables of the DTC, and $p_t$ and $p_{t-1}$ of EB. Moreover, models, which do not use the diurnal seasonal term, are referred according to their AR order. For instance AR model, which only use $p_t$ and $p_{t-1}$ as inputs, is called AR(2). VAR model using the same input variables will be known as VAR(2). In case that an AR model only uses $p_t$ as input, it is referred as AR(1).

The vector autoregressive (VAR) and vector autoregressive with exogenous input (VARX) frameworks will be compared with the AR(2) diurnal framework, both in off-line and on-line models. For instance, off-line AR are compared with off-line VAR.

The Gradient Boosting models present a similar designation. GB VAR uni, or GB VAR(2) diurnal uni, refers to a GB VAR univariate model, using $p_t$, $p_{t-1}$ and $p_{t-24+k}$, as input, while GB VAR multi, or GB VAR(2) diurnal multi, refers to a GB VAR multivariate model, using the inputs $p_t$, $p_{t-1}$ and $p_{t-24+k}$.

Table 4.1, Table 4.2 and Table 4.3 summarize the tests realized during this dissertation for both off-line and on-line models, respectively. Both on-line and off-line methods were tested for two different levels of data structures. At Photovoltaic panel (PV), it was used PV generation data series registered by Energy Box (EB). This data were aggregated according to the geographical distribution at secondary substation level. From the EB aggregation results the Distribution Transformer Controller (DTC) data series.

In off-line models, beside the distinct quantile probability, $\tau$, tested in OLS models, different shrinkage coefficients, $\nu$, and number of groups, ng, were also tested in GB.

In on-line models, different $\tau$ and $\lambda$ were tested in combination with distinct input variables. One input combination using a diurnal term and the other one without the diurnal term.

Table 4.1 – OLS off-line models tested both EB level and DTC level

| | | | $\boldsymbol{\tau}$ |
|---|---|---|---|
| **Off-line** | OLS | AR | 0.85,0.95 |
| | | VAR | 0.85,0.95 |
| | | VARX | 0.85,0.95 |

Table 4.2 – GB off-line models tested both EB level and DTC level

| | | | $\boldsymbol{\tau}$ | $\boldsymbol{\nu}$ | **ng** |
|---|---|---|---|---|---|
| **Off-line** | GB uni | VAR | 0.85,0.95 | 0.1,0.2,0.5 | 2,5 |
| | | VARX | 0.85 | 0.1,0.2,0.5 | 2,5 |
| | GB multi | VAR | 0.85,0.95 | 0.1,0.2,0.5 | 2 |
| | | VARX | 0.85 | 0.1,0.2,0.5 | 2 |

Referring to off-line models at EB level, up to two hours-ahead better results are obtained with GB VAR univariate using $\tau = 0.95$ and $\nu = 0.1$. For three and six hours-ahead better results are obtained with GB multivariate using $\tau = 0.85$ and $\nu = 0.1$. While for four and five hours-ahead better results are obtained with GB univariate using $\tau = 0.85$ and $\nu = 0.1$.

Referring to DTC level $\tau = 0.85$ performs better than $\tau = 0.95$. GB univariate VARX using $\nu = 0.2$ performs better for all horizons above one-hour-ahead. Up to one hour-ahead the best model is provided by GB univariate VAR using $\nu = 0.1$

Table 4.3 – On-line models tested both EB level and DTC level

| | | | $\tau$ | $\lambda$ |
|---|---|---|---|---|
| **On-line** | RLS | AR(2) diurnal | 0.85,0.95 | 1,0.98 |
| | | AR(2) | 0.85,0.95 | 1 |
| | | AR(1) | 0.85,0.95 | 1 |
| | | VAR(2) diurnal | 0.85,0.95 | 1,0.98 |
| | | VAR(2) | 0.85,0.95 | 0.98 |
| | | VARX(2) diurnal | 0.85,0.95 | 1,0.98 |
| | | VARX(2) | 0.85,0.95 | 1 |

On-line models best results were provided by the methods using $\lambda = 1$, both at EB level and DTC level.

At EB level the best results are provided by the VAR model using $\tau = 0.95$ and $\lambda = 1$ for all time horizons. While, at DTC level the best results are provided using $\tau = 0.85$ and $\lambda = 1$ for all time horizons. Up to one hour-ahead the best model is the VAR(2), while for the other time horizons VAR performs better

It is also important to note that VARX models using normalized EB values were tested on both OLS and RLS method. However, their results are worse than those obtained with non normalized EB values. The GB multivariate method tests using inverse covariance matrix, as gamma matrix, were run and it presented worse results than using identity matrix as gamma. Attending that this thesis aims to present the best forecasting using the proposed methods, the tests results using EB normalized, in VARX models, and inverse covariance matrix in GB multivariate function are not shown.

## 4.3 Off-line Models

### 4.3.1 OLS models

This section presents the results obtained with the off-line models. We compare the AR $nRMSE$ generated by using different quantile probability, $\tau = 0.85$ and $\tau = 0.95$, as clear sky function parameter.

Figure 4.2 and Figure 4.3 present the results obtained for AR model, which used distinct $\tau$ parameters to obtain the clear sky values at DTC and EB levels, respectively.

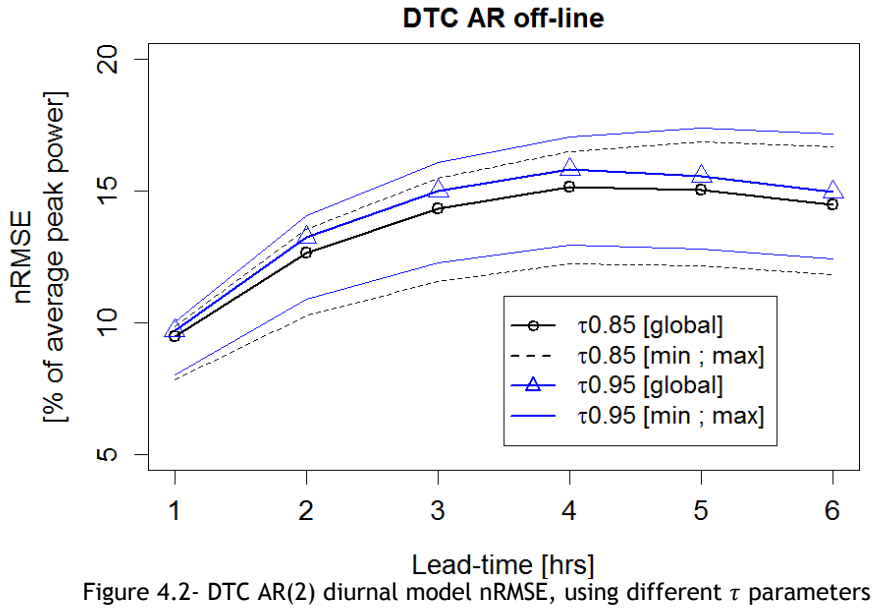


Figure 4.2- DTC AR(2) diurnal model nRMSE, using different $\tau$ parameters

From analyzing Figure 4.2 and Figure 4.3, it is concluded that the AR model provides higher-accuracy for both DTC and EB with $\tau$ equal to 0.85.

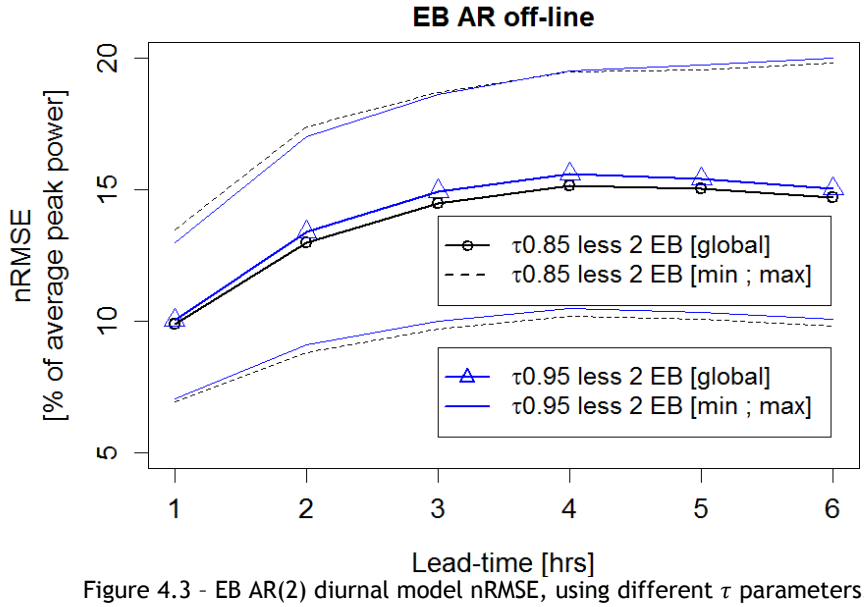


Figure 4.3 – EB AR(2) diurnal model nRMSE, using different $\tau$ parameters.

Note that two EB were not included in the results of Figure 4.3 since the nRMSE values were much different from the others.

The two EB excluded present short time periods without measurements. In addition, these measurements at the beginning of time series are not consistent with those from the time series.

If the two EB were included in Figure 4.3, the maximum nRMSE for each time horizon would vary between 17.0%, for one hour-ahead up to 30.0% for three hours-ahead. Without considering the two EB and using $\tau = 0.85$, the maximum nRMSE values range from 13.5%, for one hour-ahead, up to 19.8%, for six hours-ahead $\tau = 0.85$. However, considering all EB with $\tau = 0.95$, the maximum nRMSE varies between 15.5% for one hour-ahead, and 23.5% for three hours-

ahead. And excluding the two EB, the maximum nRMSE range from 13.0% for one hour-ahead up to 20.0% for six hours-ahead.

Analyzing $\tau = 0.85$ mean values and excluding the two EB, the values varies from 9.6% to 14.6%; while considering all EB, the maximum nRMSE range from 9.8% to 15.3%. Using $\tau = 0.95$ values, the fluctuation is lower. While excluding two EB, the maximum nRMSE presents values between 9.7% and 15.1%, values range from 9.9% and 15.3% were obtained when all EB were considered.

Attending that both plots (Figure 4.2 and Figure 4.3) have the same y-axis scale, we can inferred that EB forecasting produces less accurate results and more dispersed than the DTC forecasting. This conclusion can be explained by the fact that the values of the DTC are the aggregate of various EB and the aggregated forecast errors tend to cancel between them.

DTC average $nRMSE$ are smaller than EB average $nRMSE$ by 0.7% for six hours-ahead and 1.2% for three hour-ahead when $\tau = 0.85$is considered. However, for $\tau = 0.95$, DTC results are better in a range of 0.33%, for six hours-ahead and 0.77%, for one hour-ahead.

From these analyses, $\tau = 0.85$ will be used as the AR model for the other off-line models.

Analyzing the nBias at DTC, calculated using Eq. (4.3), one can realize that the predictions overestimate the value, i.e., negative $nBias$, for all DTC at all time horizon.

Figure 4.4 shows the nBias results at DTC and EB levels, using $\tau = 0.85$. Notwithstanding that few EB show positive $nBias$, the negative signal is almost the general trend. In fact, the average of EB normalized mean error is negative for all time horizons.

Moreover, just 3 out of 44 values (6.8%) present positive $nBias$ for one, two, five and six hours-ahead time horizons, while for three and four hours-ahead time horizons there are 4 EB values (9.1%) presenting positive $nBias$.

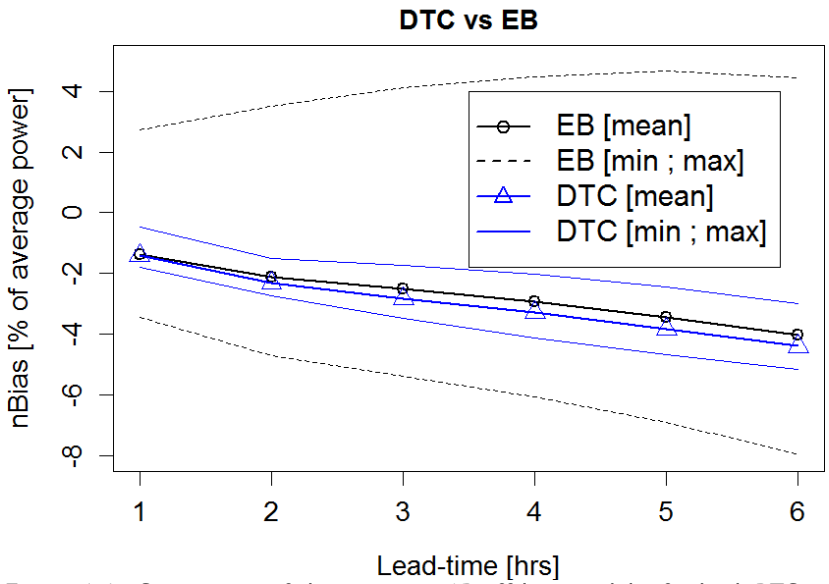


Figure 4.4 - Compraison of the $nBias$, in AR off-line models, for both DTC and EB

Figure 4.5 shows the improvement over AR obtained with VAR, at both DTC and EB level, and obtained with VARX, at DTC level.

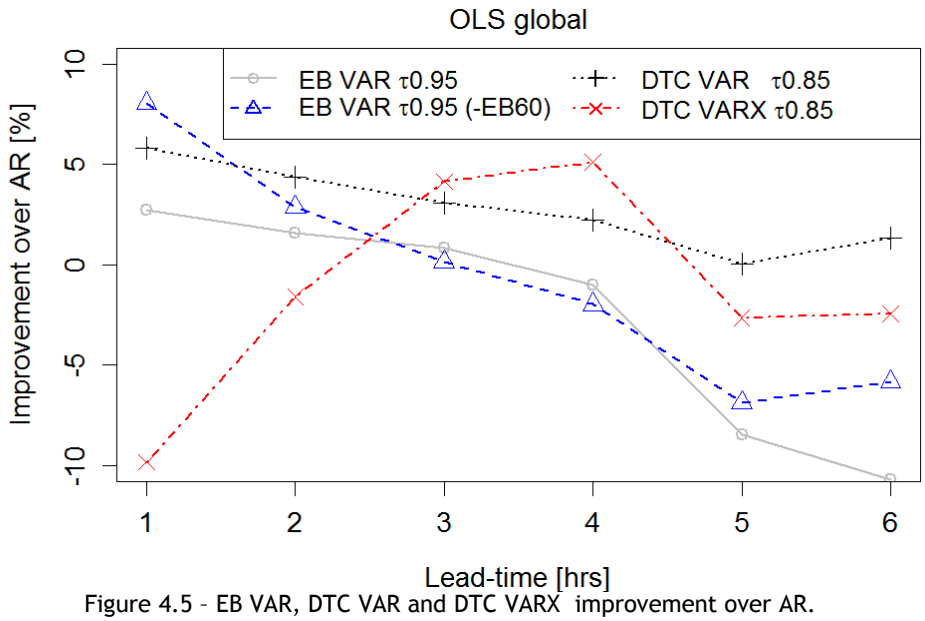


Figure 4.5 – EB VAR, DTC VAR and DTC VARX improvement over AR.

Attending that the area, in which the panels are scattered, is smaller than the area covered by cloud motion in few hours, one would expect that the influence of neighborhood information in forecasting decrease as time horizon increase. Figure 4.5 supports this idea, since the VAR DTC improvement declines with time-horizon. Moreover, the VARX results approach to VAR as time-horizon increases.

In what concerns to EB, the nRMSE of EB60 was among the lower values for AR, in VAR model their nRMSE was among the higher ones. Therefore, EB60 was excluded.

Analyzing Figure 4.5 conveys that OLS have problems dealing with high dimension cases, and the VAR at DTC level results are more stable and generally better than VARX at DTC level and VAR at EB level. Note that while VAR for DTC is a 31-predictors model, VARX is a 119-predictors model and VAR at EB level is a 133-predictors model.

The global DTC VAR improvement over AR $\tau = 0.85$ varies between 5.83% for one hour-ahead, and 0.04% for five hours-ahead, when $\tau = 0.85$ is used. Moreover, all DTC present positive improvement for one, two and six hours-ahead. Referring to third and fourth lagged hours, just one of 10 DTC in each time-horizon presents negative improvement. For five hours-ahead, three DTC present negative improvements but the values are lower than 1% in two of them.

While the higher individual improvement, 9.85%, is presented for one hour-ahead, the worst value, -4.72%, appears at five hours-ahead.

The variability on the VARX results make the analysis will be difficult. Table 2.1 presents the maximum (Max) and minimum (min) improvements achieved by the VARX for one DTC and for each time horizon.

Table 4.4 – Improvement OLS VARX over AR (%)

| | $(t+1)$ | $(t+2)$ | $(t+3)$ | $(t+4)$ | $(t+5)$ | $(t+6)$ |
|---|---|---|---|---|---|---|
| **Max (%)** | 5.81 | 8.51 | 10.81 | 10.89 | 5.13 | 3.68 |
| **Min (%)** | -26.61 | -13.91 | 1.87 | 1.24 | -10.06 | -6.88 |

EB values, which are not normalized when used as exogenous input variables, may justify the VARX inconsistency. Normalized EB values, using clear sky function as exogenous variables, were tested. However, they present worst results than the model with non-normalized values.

Figure 4.6 presents the excluded EB measure time series and one can realize the changes that occur in PV generation of this EB across time. This EB nRMSE is among the lower values for AR(2) diurnal of-line model indicating that the changes do not affect the relation between the input variables and the generation values. This behavior is coherent with installed capacity lost. For instance, if the house where this EB is connected loses part of the installed capacity, the generation values are modified but the relation between output and input variables will not be changed. This fact explains the AR results. However, as VAR model uses the neighborhood information, the relation between the outputs of the EB, which change its generation, and the input of the others modifies. This explains the results of this EB in VAR model.

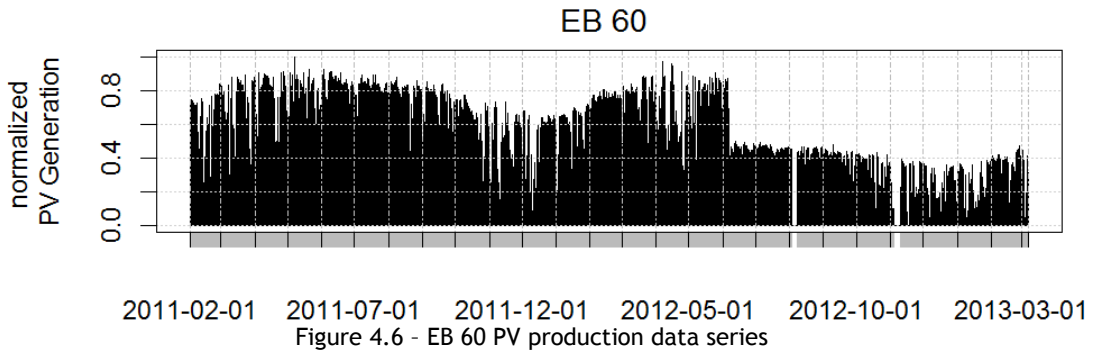

Figure 4.6 – EB 60 PV production data series

The predictions obtained using the VAR model, as shown in Figure 4.7, tend to be overestimated for the first two lag hours and underestimated from four to six hours-ahead.

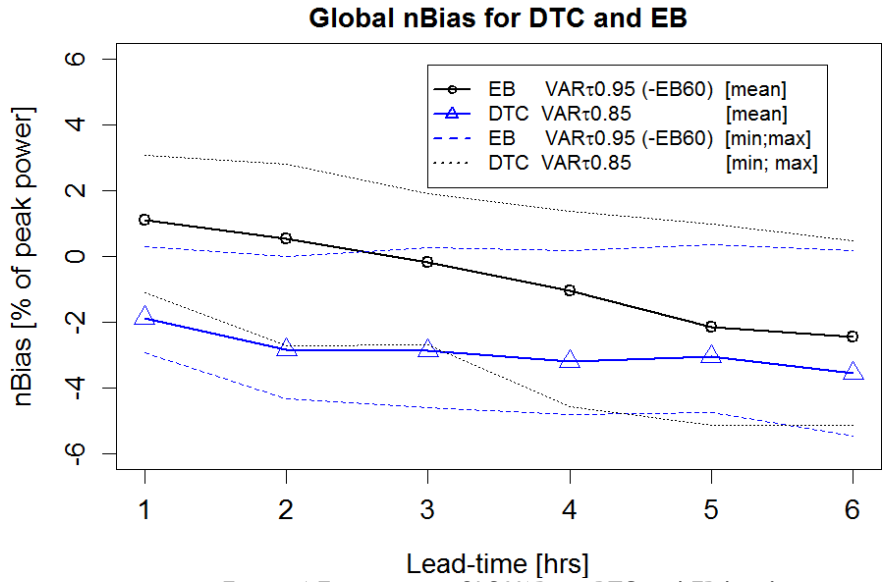

Figure 4.7 - $nBias$, in OLS VAR, at DTC and EB level

Furthermore, Table 4.5 shows that a considerable number of the EB presented, for each time-horizon, the same $nBias$ signal.

Table 4.5 - Number of EB and the percent of $\tau = 0.95$ (-EB 60) which present negative $nBias$

| | $(t+1)$ | $(t+2)$ | $(t+3)$ | $(t+4)$ | $(t+5)$ | $(t+6)$ |
|---|---|---|---|---|---|---|
| **EB** | 3 | 10 | 22 | 33 | 40 | 42 |
| **EB (%)** | 6.98 | 23.26 | 51.16 | 76.74 | 93.02 | 97.67 |

### 4.3.2 Gradient-Boosting univariate

This section presents the GB VAR univariate method results. GB VAR univariate forecasts individually each EB or DTC. Thus, GB univariate function determines:

- 31 predictors at once in GB VAR at DTC level
- 119 predictors at once in GB VARX
- 133 predictors at once in GB VAR at EB level.

Note that GB should select the variables that produce the most accurate forecasting model, thus one expects that some predictors coefficients became zero.

GB functions, used in this dissertation, present as parameters the shrinkage parameter, $\nu$, and the number of iterations, mstop. According the literature, $\nu$ should lies in the interval $]0; 1]$, thus, both GB univariate and GB multivariate tests run using $\nu\{0.1,0.2,0.5,1\}$.

As mentioned, GB functions have two important parameters, which are the shrinkage coefficient, $\nu$, which is an empirical parameter that is determined by trial-error experiences, and the number of iteration, mstop. To determine mstop, GB can resort to CV function using a high number of maximum iterations (Mmax). CV function selects the optimal mstop from Mmax reaching the better results.

Cross validation can determines mstop, using as parameters the maximum number of iterations (Mmax) and the number of groups in which train data series are divided (ng).

Thus, Figure 4.8 provides an overview of the evolution of the square error for one hour-ahead and for several EB. Analyzing the figure it is clear that the error rapidly decreases and stabilizes as the iteration rises. In some cases, the error trend to rise after reaching the minimum such as in EB 41. However, the figure scale does not allow us to realize this behavior.

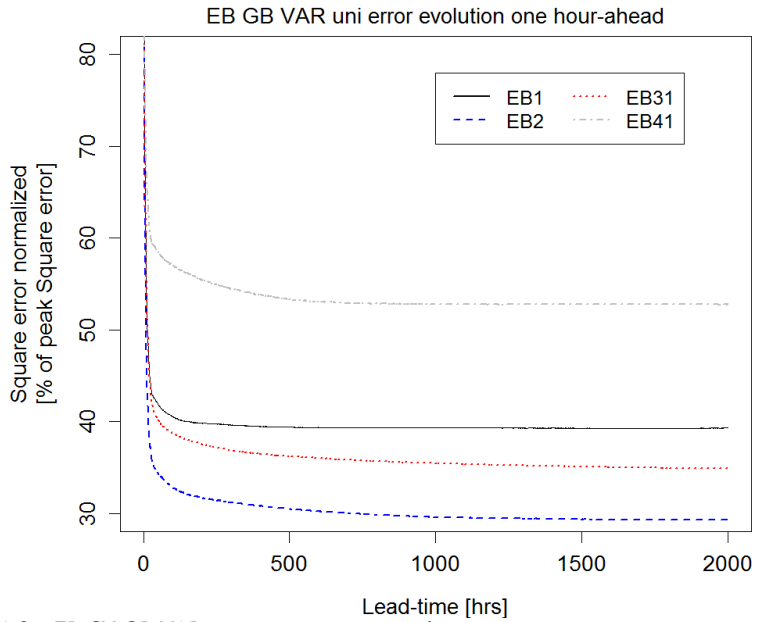


Figure 4.8 - EB CV GB VAR uni square error evolution using $\nu = 0.1$, $\tau = 0.85$ and ng=2

The error rising trend is clear in six hours-ahead results.

Figure 4.9 presents an analog study up to to six hours-ahead. In this case, is evident that the error trend to rise after reaches a minimum.

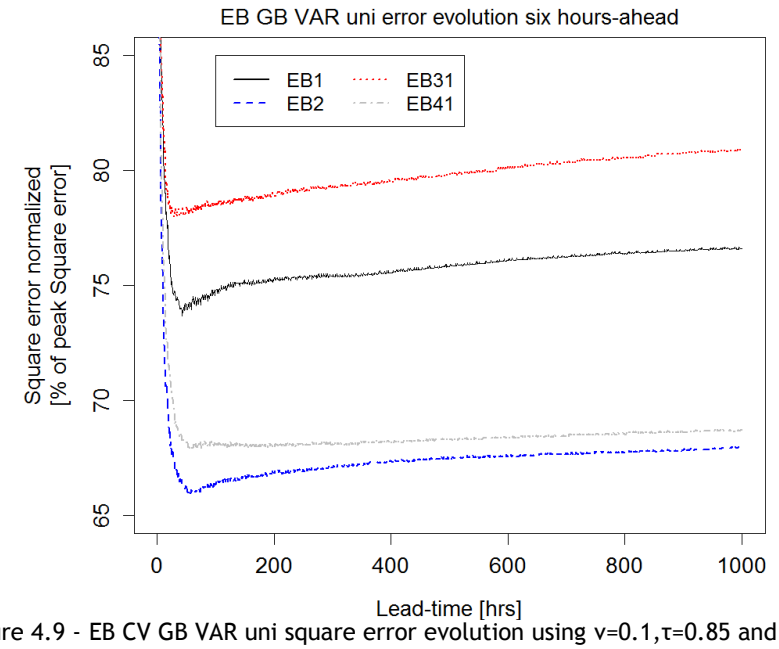


Figure 4.9 - EB CV GB VAR uni square error evolution using ν=0.1,τ=0.85 and ng=2

Figure 4.10 presents the improvement up to six hours-ahead obtained with GB univariate in cross validation function and using distinct number of groups (ng) and distinct shrinkage coefficients.

Analyzing Figure 4.10, one can realize that $\nu$ is an important parameter in GB VAR univariate model performance. In cross validation function also ng has influence in results. The combination that provides better results, $\nu = 0.1$ and ng=2, is that one which presents less computational effort.

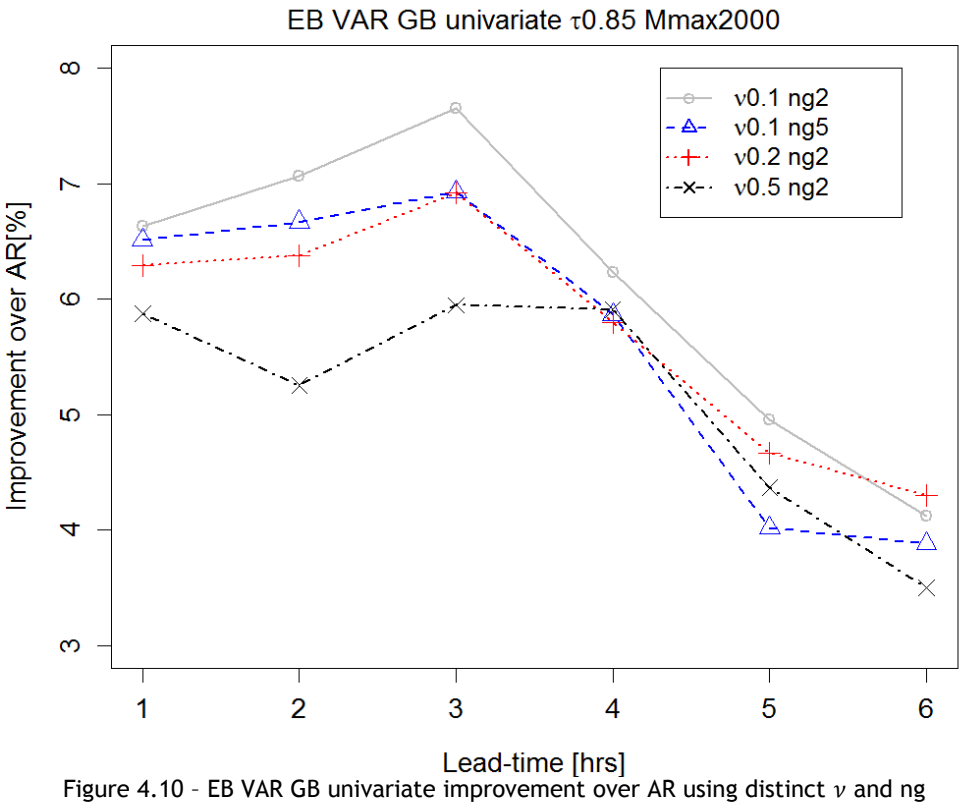


Figure 4.10 – EB VAR GB univariate improvement over AR using distinct $\nu$ and ng

Attending that using the quantile probability, $\tau$, 0.95, best results achieved with $\nu = 0.1$, Figure 4.11 presents a results comparison between $\tau = 0.85$ and $\tau = 0.95$. As figure shows the best results achieve for one and two hours-ahead using $\tau = 0.95$, while for three to six hours-ahead $\tau = 0.85$ performs better.

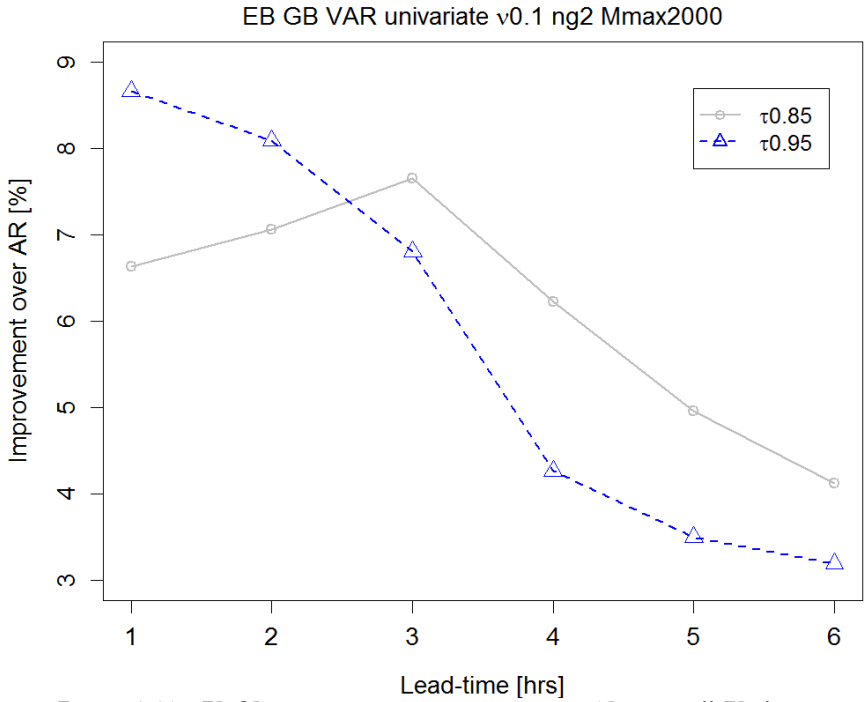


Figure 4.11 - EB GB univariate improvement over AR using all EB dataset

Figure 4.12 shows the EB GB uni average nBias comparing different $\tau$, $\nu$.

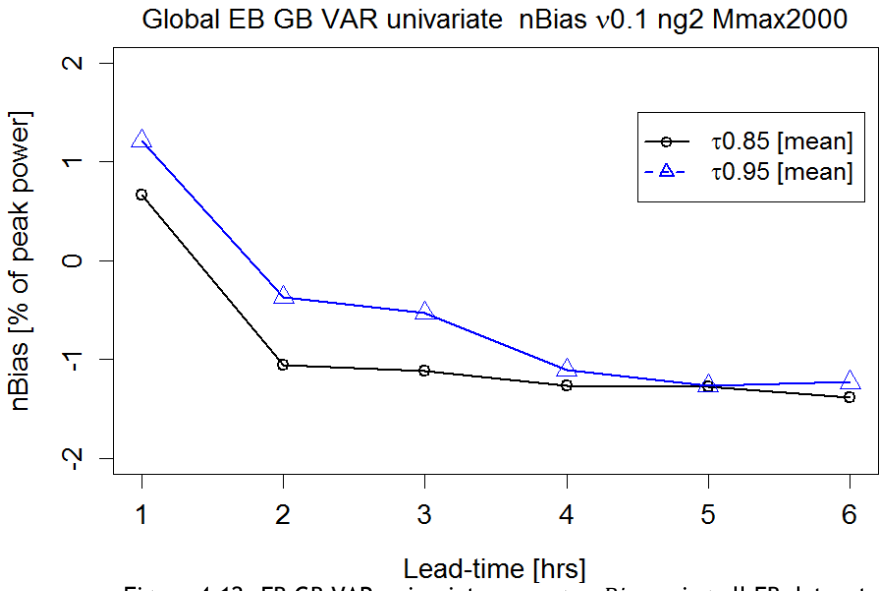


Figure 4.12- EB GB VAR univariate average $nBias$ using all EB dataset

The $nBias$ trend is clear, for all time horizons, since that an important percentage of EB present the same nBias signal for each time-horizon. Taking GB VAR(2) diurnal uni model using $\nu = 0.1$ and $\tau = 0.95$, for one hour-ahead, just 15.9% of EB present negative $nBias$. While, GB VAR(2) diurnal uni model using $\nu = 0.1$ and $\tau = 0.85$ present, from two to six hours-ahead, negative nBias signal between 86.4%, for three hours-ahead, and 90.9%, for two and six hours-ahead. Thus, one can conclude that, for one hour-ahead, GB univariate trend to overestimate, while for others time horizons trend to underestimate.

GB ability to select the inputs variables that provide more information to the forecasting system supported the idea of use GB.

The predictors matrix of GB univariate using $\nu = 0.1$ and $\tau = 0.85$, for one hour-ahead, shows that, in average terms, the most important information came from the $p_t$ variable, including its own $p_t$ variable in 40 of the 44 EB. The coefficients related with this information are non-zero, in average terms, 49%.

As mentioned in section 3.2, the clear sky function does not remove all seasonality. Attending that, decided include $p_{t-1}$ as input variable, which is justified the 33% non-zero $p_{t-1}$ coefficient predictors.

In what concerns to predictor coefficients related to the day before input 23% are non-zero. Related to this variable 40 EB present non-zero coefficient in its own $p_{t-24+k}$ input, pointing that model is uses this information to compensate seasonality.

Providing a similar analysis, for six hours-ahead, resorting GB uni using $\nu = 0.1$ and $\tau = 0.85$ it is clearly that predictors matrix is sparser. While, in one hour-ahead, 3497 of 5808 (60%) of the predictors coefficients are zero, in six hours-ahead, 5025 of the 5808 (87%) coefficients are zero.

In average terms, five $p_{t,i}$, two $p_{t-1,i}$ and eleven $p_{t-24+k,i}$ coefficients of each EB are non-zero. This numbers reflect the importance of diurnal term in longer time-horizons.

The previous analysis shows that GB univariate, as expected, it is capable of leading with high dimensional problem. However, attending that no EB can be ruled out from the models, which implies that all the information must be send for a central point, do not provide a solution for the communication problem.

Figure 4.13 shows the performances, at DTC level, up to six hours-ahead, obtained using $\tau = 0.85$. As verified at EB level, at DTC level both VAR and VARX provided better results using small $\nu$ values, reader can find more details in Appendix. The VARX improvement over VAR, at DTC level, is positive beyond one hour-ahead supporting the idea that the EB data series provide information about clouds behavior.

Although the improvement achieved by the GB VARX uni points out that forecasting for two and three hours-ahead are better than for one hour-ahead, this idea is wrong. The nRMSE for two and three hours-ahead is greater than for one hour-ahead when GB VARX uni model is considered. However, as the nRMSE increases slower in GB VARX uni model than in AR model, the improvement over AR is higher even when the GB VARX uni model nRMSE increases.

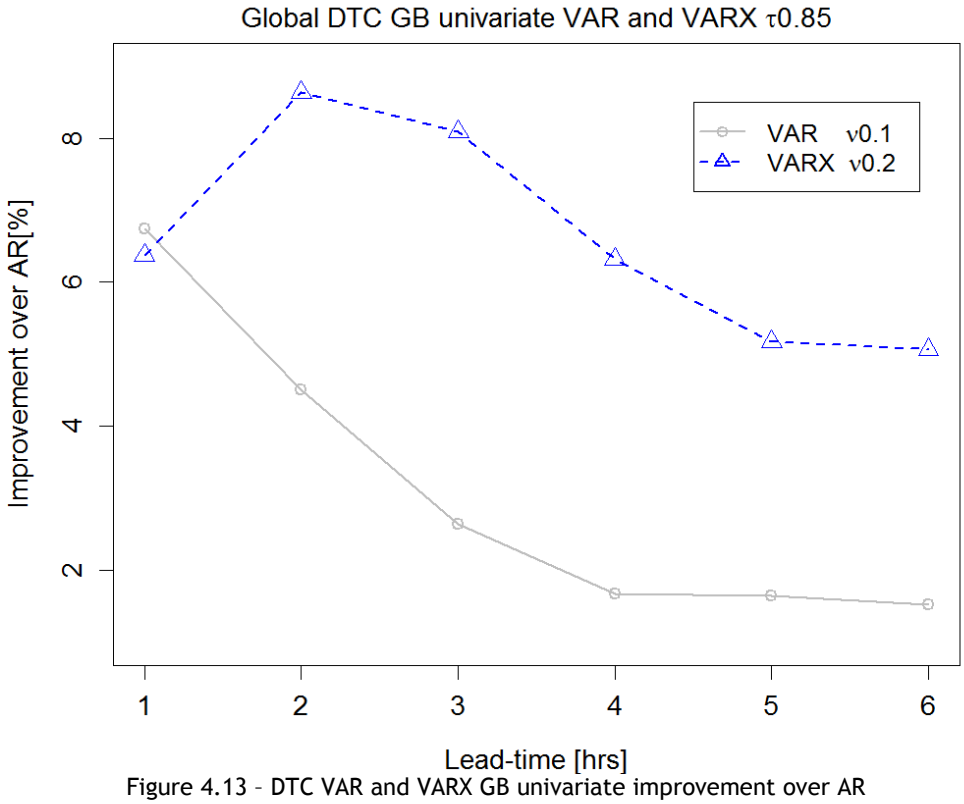


Figure 4.13 - DTC VAR and VARX GB univariate improvement over AR

Table 4.6 presents the number of iterations determined by CV to DTC GB uni, showing that all DTC present mstop smaller than 1100, with exception of DTC5.

Table 4.6 - DTC GB VAR uni mstop obtained with CV using $\tau = 0.85$ and Mmax=2000

| | $DTC1$ | $DTC2$ | $DTC3$ | $DTC4$ | $DTC5$ | $DTC6$ | $DTC7$ | $DTC8$ | $DTC9$ | $DTC10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $(t+1)$ | 248 | 301 | 949 | 544 | 1999 | 752 | 1077 | 364 | 496 | 494 |
| $(t+2)$ | 223 | 250 | 931 | 344 | 1894 | 312 | 862 | 290 | 277 | 461 |
| $(t+3)$ | 223 | 185 | 157 | 373 | 1997 | 115 | 450 | 233 | 181 | 221 |
| $(t+4)$ | 197 | 142 | 143 | 113 | 1830 | 114 | 95 | 170 | 157 | 217 |
| $(t+5)$ | 148 | 118 | 95 | 110 | 1231 | 113 | 109 | 102 | 124 | 165 |
| $(t+6)$ | 145 | 76 | 69 | 80 | 1002 | 77 | 81 | 83 | 110 | 121 |

As GB VAR uni at DTC level is a small dimension problem, it is not expected that the predictors matrix is as sparse as in GB VAR uni at EB level or GB VARX uni. Checking the coefficient matrix for one hour-ahead and using GB univariate model with $\tau = 0.85$ and $\nu = 0.1$, 93 of 300 (31%) of the coefficients are zero. In DTC predictors matrix, every $p_t$ coefficients referring to its own DTC are non-zero. In global terms, 22% of $p_t$referring predictors are zero, while 28% of $p_{t-1}$ coefficients are zero and 43% of $p_{t-24+k}$ coefficients are zero. Referring to six hours-ahead, the number of zero $p_{t-24+k}$ coefficients is 61 of 100 (61%). However, all $p_{t-24+k}$ coefficients, referring to its own DTC, are zero. The number of $p_t$ coefficients that are zero is 62, and the number of $p_{t-1}$coefficients equals to zero is 82.

GB VARX univariate model predictors matrix presents high sparsity, mostly in predictors related with EB. Considering VARX GB univariate model with $\tau = 0.85$ and $\nu = 0.1$, the sparsity related to EB is higher than related to DTC. While for the predictors related with DTC, 33% of $p_{t,DTC}$, 21% of $p_{t-1,DTC}$ and 42% of $p_{t-24+k,DTC}$ are zero, for the predictors related with EB, 72% of $p_{t,EB}$ and 73.4 % of $p_{t-1,EB}$ are zero.

However, Table 4.7 – DTC GB VARX uni mstop obtained with CV using Mmax=2500 indicates that the error in VARX GB univariate model with $\tau = 0.85$ and $\nu = 0.1$ will decrease when more iterations are performed. Given that the mstop is closer to 2500, except for six hours-ahead, it is shown that error is in a descendent trajectory.

Table 4.7 – DTC GB VARX uni mstop obtained with CV using Mmax=2500

| | $DTC1$ | $DTC2$ | $DTC3$ | $DTC4$ | $DTC5$ | $DTC6$ | $DTC7$ | $DTC8$ | $DTC9$ | $DTC10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $(t+1)$ | 2500 | 2494 | 2483 | 2454 | 2487 | 2500 | 2498 | 2466 | 2500 | 2464 |
| $(t+2)$ | 2492 | 2494 | 2497 | 2295 | 2491 | 2495 | 2496 | 2462 | 2500 | 2497 |
| $(t+3)$ | 2500 | 2460 | 2484 | 2438 | 2500 | 1824 | 2495 | 2364 | 2496 | 2499 |
| $(t+4)$ | 2492 | 2462 | 2445 | 2483 | 2494 | 2047 | 1373 | 2492 | 2483 | 2498 |
| $(t+5)$ | 2498 | 2500 | 2470 | 2499 | 2498 | 2498 | 2474 | 2500 | 2484 | 2468 |
| $(t+6)$ | 2462 | 477 | 2483 | 2493 | 2422 | 1348 | 1424 | 755 | 1259 | 757 |

### 4.3.3 Gradient-Boosting multivariate

This section devotes to the analysis of the GB multivariate results. Like in previous section, we present the EB GB VAR multivariate results, including improvement and predictors matrix sparsity analysis, and we provide a similar analysis for both DTC GB VAR multivariate and DTC GB VARX multivariate models.

The GB multivariate function requires a gamma matrix that relates several units to including in the forecasting. We have tested two distinct matrices as gamma: identity matrix and error covariance matrix using the GB univariate results. As the results using the identity matrix as gamma were better, this work focus on the results obtained by using covariance matrix as gamma.

The GB univariate function calculates for each EB or DTC the predictor coefficients. Then, in each iteration, one among 133 predictors, one among 119 and one among 11 are selected and changed in the EB case, in the DTC VARX version and in DTC VAR version, respectively. On the other hand, the whole dataset is considered at the same time in the GB multivariate version, which increases the problem dimension. In each GB iteration, for instance on EB problem, one among 5852 predictor coefficient is just adjusted. Thus, it is expected that the

number of iterations in the GB VAR multivariate function providing the best results became greater than in GB univariate function.

Figure 4.14 and Figure 4.15 show EB GB multivariate improvement over OLS AR (AR) using different $\nu$, for $\tau = 0.85$ and $\tau = 0.95$, respectively.

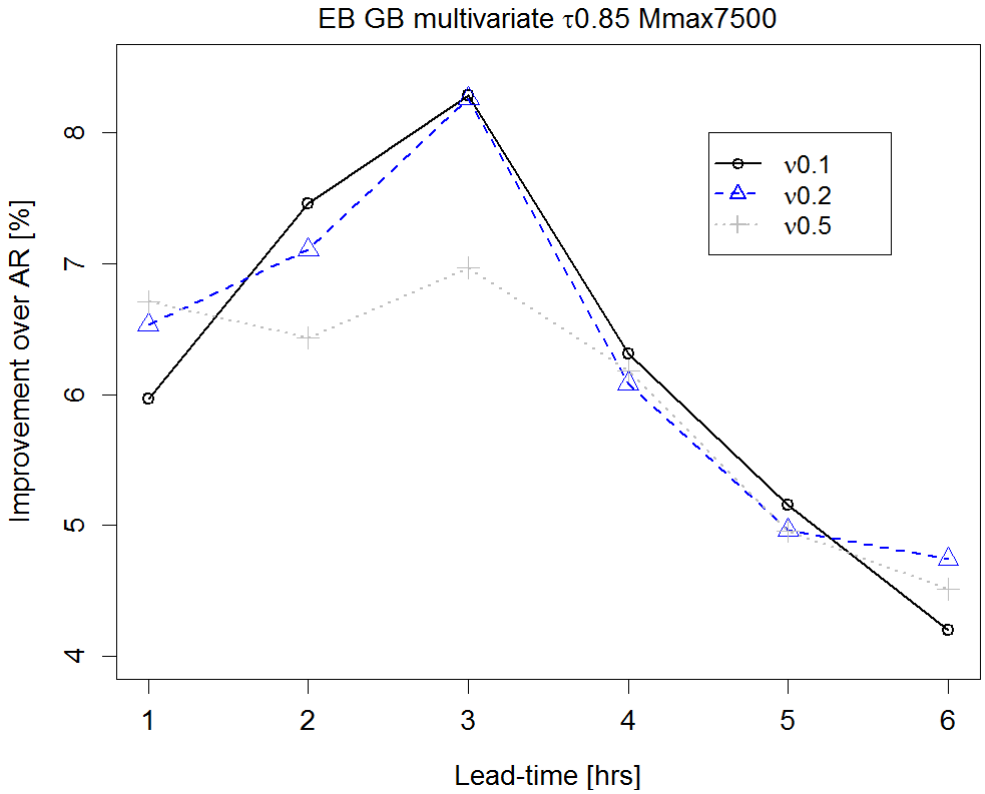


Figure 4.14 - EB GB VAR multi imp. over AR comparing different $\nu$ , with $\tau = 0.85$

Like in GB VARX uni, the fact that the error still decreases after 7500 iterations leads to obtain the best results with higher $\nu$. The results obtained with GB multivariate are very similar to GB univariate results. The improvement for two hours-ahead is higher than for one hour-ahead, and $\tau = 0.95$ presents better results up to two hours-ahead. However, $\tau = 0.95$and $\nu = 0.5$ provide the best results up to three hours-ahead.

Analyzing the GB VAR multivariate predictors matrix obtained using $\tau = 0.85$ and $\nu = 0.1$as model parameters for one hour-ahead, the matrix sparsity is obvious and 80.0% of the predictors coefficients are zero. The model for one hour-ahead uses 502 of 1936 (25.9%) $p_t$ related predictors, 344 of 1936 (17.8%) $p_{t-1}$ related predictors and 274 (14.2%) $p_{t-24+k}$ related predictors. The EB7, EB8, EB43 and EB60 $p_t$ related coefficients are zero in all EB meaning that the information from these EB is not important for the forecasting system.

The DTC GB VAR multivariate model predictors coefficients matrix, where $\tau = 0.85$, $\nu = 0.1$ and mstop=1000 are used and for one hour-ahead, is composed by 53.3% of zero coefficients, $p_t$with 36%, 54% and 70% of null coefficients for $p_t$, $p_{t-1}$ and $p_{t-24+k}$, respectively. $p_{t-1}$ $p_{t-24+k}$For six hours-ahead, the overall zero coefficients is 68%, distributed as 51% for $p_{t-24+k}$, 64% for $p_{t-1}$ and 89% for $p_{t-1}$.

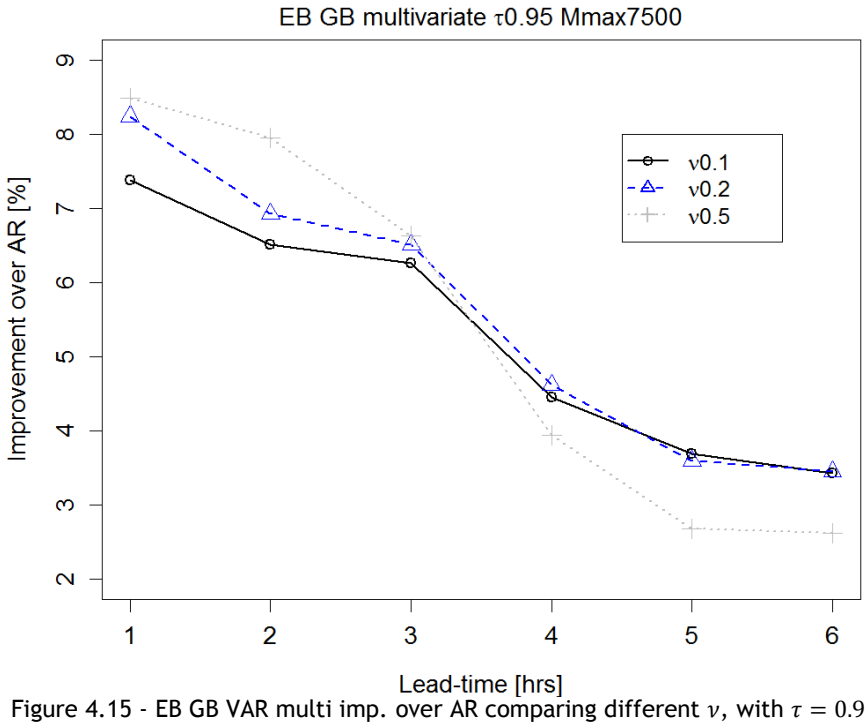


Figure 4.15 - EB GB VAR multi imp. over AR comparing different $\nu$, with $\tau = 0.95$

Similar to the behavior of the DTC GB univariate model, Figure 4.16 shows that DTC GB VARX performs better than DTC GB VAR above one hour-ahead. This fact emphasizes the cloud motion information provided by EB.

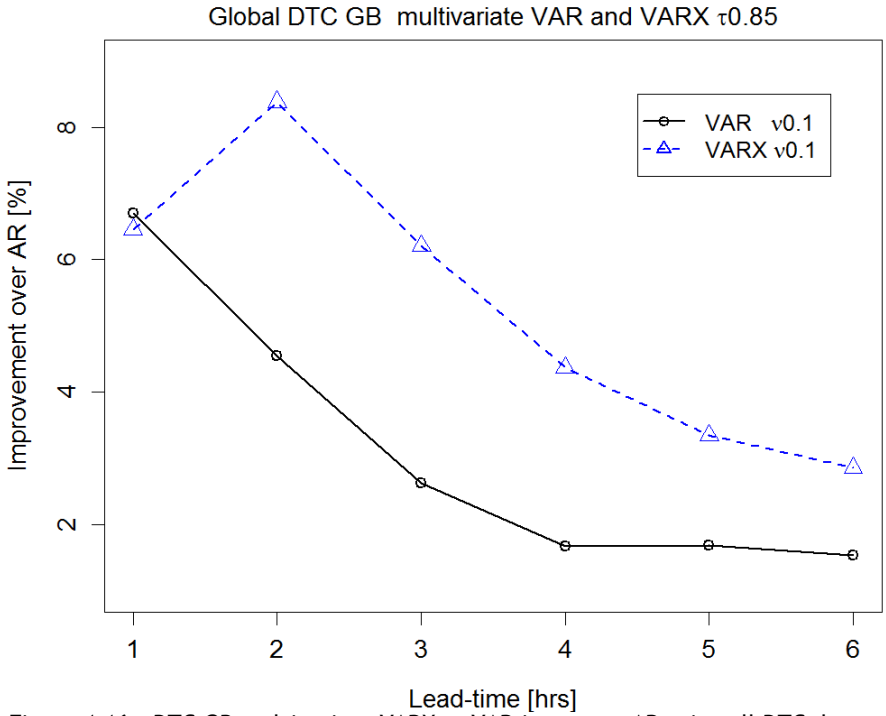


Figure 4.16 - DTC GB multivariate VARX vs VAR imp. over AR using all DTC datasets

### 4.3.4 GB and OLS comparison

In this section, we compare the OLS, GB univariate and GB multivariate models. The main aim is to summarize the off-line results and emphasize the best results.

Figure 4.17, compares GB uni, GB multi and OLS, showing how GB can deal with EB problem dimension.

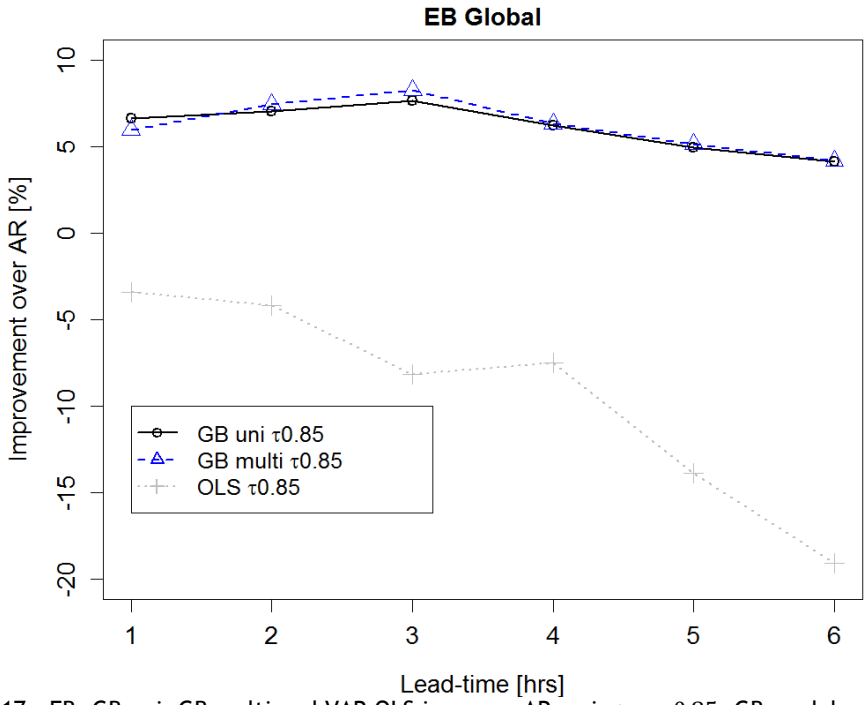


Figure 4.17 - EB, GB uni, GB multi and VAR OLS imp over AR, using $\tau = 0.85$. GB models use $\nu = 0.1$

Figure 4.17 also confirms that the GB VAR both univariate and multivariate provide similar results when the right number of iterations are used. However, the computational time cost is different. While, GB uni function with mstop=500 takes 8 minutes and 35 seconds to forecast the 44 EB up to six hours-ahead, the GB multi function with mstop=7500 takes 23 minutes and 35 seconds. This results were obtained in an Intel Core i7-2600 CPU 3.40GHz with 8.00GB of RAM.

Although the global results achieved by both GB uni and GB multi models are similar, the individual forecasts are distinct. For instance, EB 12 GB uni is 5.5% better than EB 12 GB multi for one hour-ahead, and EB 17 GB multi is 0.36% better than GB uni. However, the opposite scenario is shown for six hours-ahead; GB multi can be 4.6% better than GB uni, and GB uni is only 0.8% better than GB multi.

The improvement of both univariate and multivariate GB VAR over the AR model is presented in Figure 4.18 and it can reach up to 45% for three hours-ahead. Note that the higher improvement, verified between two and six hours-ahead, is not caused by the decrease of the error in VAR models for these horizons, but rather the error growth slower compared to the AR model.

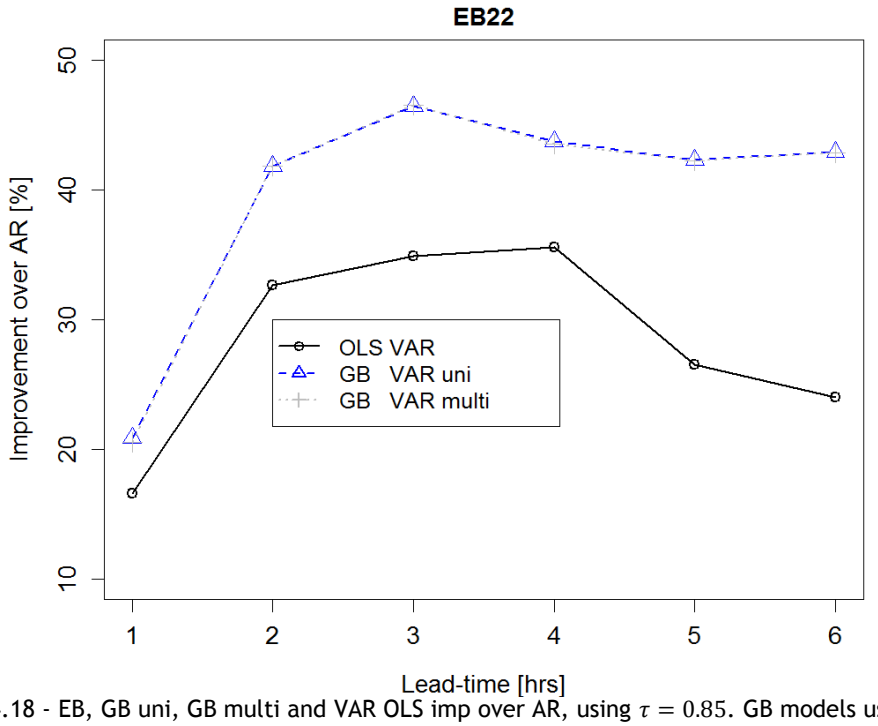


Figure 4.18 - EB, GB uni, GB multi and VAR OLS imp over AR, using $\tau = 0.85$. GB models use $\nu = 0.1$

It is interesting to note that the differences between VAR OLS and GB VAR are improved in both univariate and multivariate models (shown in Figure 4.19) and this difference can be as large as 40%.

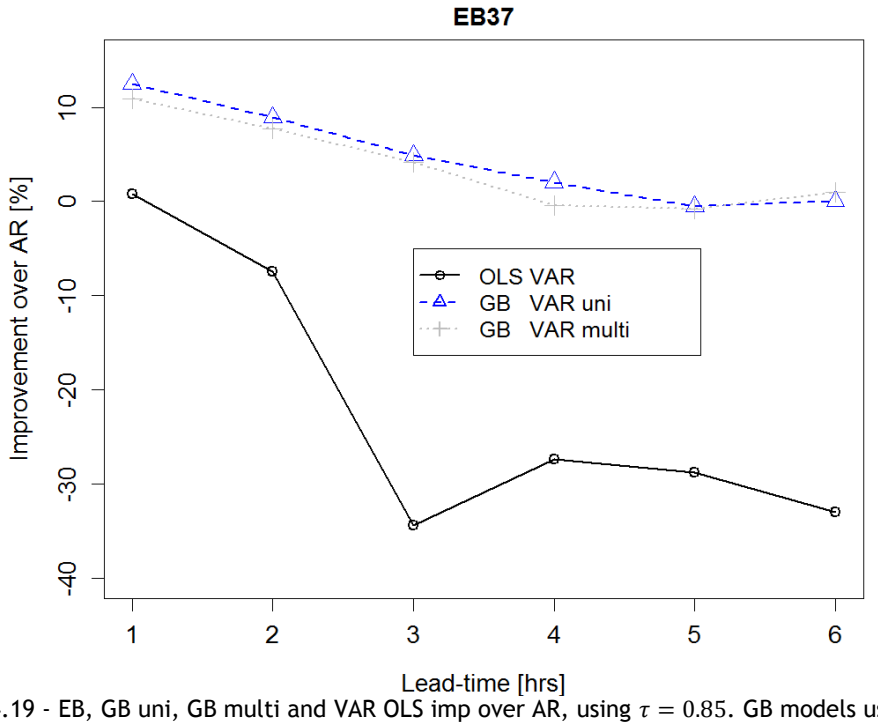


Figure 4.19 - EB, GB uni, GB multi and VAR OLS imp over AR, using $\tau = 0.85$. GB models use $\nu = 0.1$

Figure 4.20 presents a comparison between GB, both univariate and multivariate, and the OLS for both VAR and VARX models. As in EB, the results of GB are similar in both versions (univariate and multivariate). Maybe, DTC OLS VAR model achieves results very similar to GB models due to the dimensional problems.

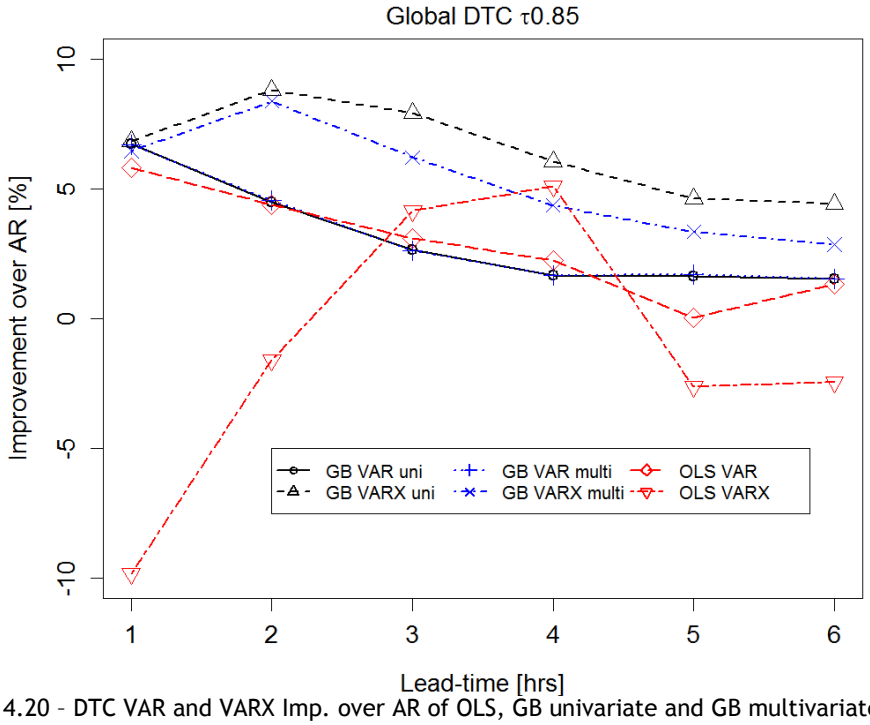


Figure 4.20 – DTC VAR and VARX Imp. over AR of OLS, GB univariate and GB multivariate models

Figure 4.21 shows a comparison between GB, both univariate and multivariate, and the OLS, for both VAR and VARX models. Note that the improvement archives 15%.

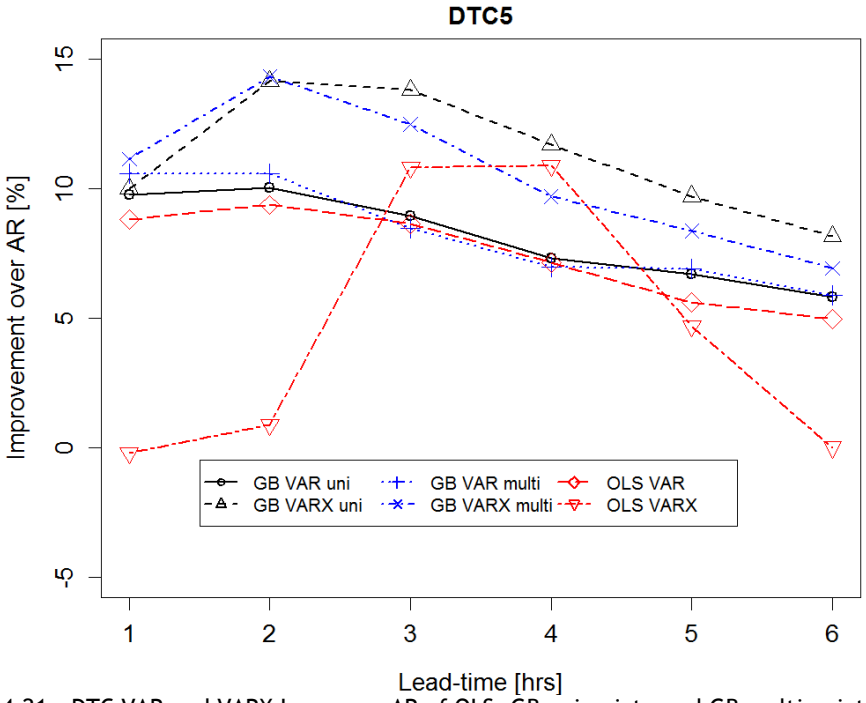


Figure 4.21 - DTC VAR and VARX Imp. over AR of OLS, GB univariate and GB multivariate models

## 4.4 On-line Models

This section analyzes the results of the adaptive recursive least squares based models (RLS). Firstly, we analyze the AR models results for both DTC and EB and they are compared with persistence. Following, we compare the VAR and VARX results with AR.

Throughout this section, different AR and VAR models have been used. AR and VAR models, where $p_t$, $p_{t-1}$ and $p_{t-24+k}$ have been considered as input, are referred as AR(2) diurnal and VAR(2) diurnal, respectively. For simplicity, we also designate these models as AR and VAR models. However, in case that both $p_t$ and $p_{t-1}$ are the models input, and $p_{t-24+k}$ is not considered, the models are referred as AR(2) and VAR(2). Concerning the VARX models, when the model inputs are DTC $p_t$, $p_{t-1}$ and $p_{t-24+k}$ and EB $p_t$, $p_{t-1}$, they are designated as VARX(2) diurnal or VARX; in case that DTC $p_{t-24+k}$ is not considered as input, the model is known as VARX(2).

In the adaptive model based on recursive least squares, it is possible to use a *forgetting factor* $(\lambda)$. The factor $\lambda$ allows reducing the earliest information weight. Given that the PV panels conditions may change in time, the objective of the factor $\lambda$ is the reduction of the influence in forecasting system of the data obtained at different operation conditions.

In all four $(\tau, \lambda)$ combinations tested, for all time horizons and all DTC, the $nBias$ is negative. Moreover, as Table 4.8 shows, EB present a similar behavior reinforcing the fact that the AR models tend to underestimate.

Table 4.8 – Number of EB, in a total of 44, which present negative $nBias$, using distinct $\tau$ and $\lambda$

| | $(t+1)$ | $(t+2)$ | $(t+3)$ | $(t+4)$ | $(t+5)$ | $(t+6)$ |
|---|---|---|---|---|---|---|
| $\boldsymbol{\tau = 0.85\ \lambda = 1}$ | 41 | 42 | 41 | 42 | 43 | 43 |
| $\boldsymbol{\tau = 0.85\ \lambda = 0.98}$ | 43 | 44 | 44 | 44 | 43 | 44 |
| $\boldsymbol{\tau = 0.95\ \lambda = 1}$ | 43 | 43 | 43 | 43 | 43 | 43 |
| $\boldsymbol{\tau = 0.95\ \lambda = 0.98}$ | 44 | 44 | 44 | 44 | 44 | 44 |

Figure 4.22, Figure 4.23 and Figure 4.24 show the forecasting behavior for a few days, chosen from the test data set.

Figure 4.22 and Figure 4.23 present a sequence of few clear sky days in a row. However, while Figure 4.22 shows days from beginning of the data set, Figure 4.23 present values from the end of the test data set. Note that, while in Figure 4.22 the underestimation of both AR and VAR models is clear, in Figure 4.23, this behavior is not obvious in the VAR model. These results point out the adaptive capability of the VAR model.

Although, the values presented in Figure 4.22 and Figure 4.23 are different, the improvement of almost 25%, shown in Figure 4.34, for one hour-ahead is not justified.

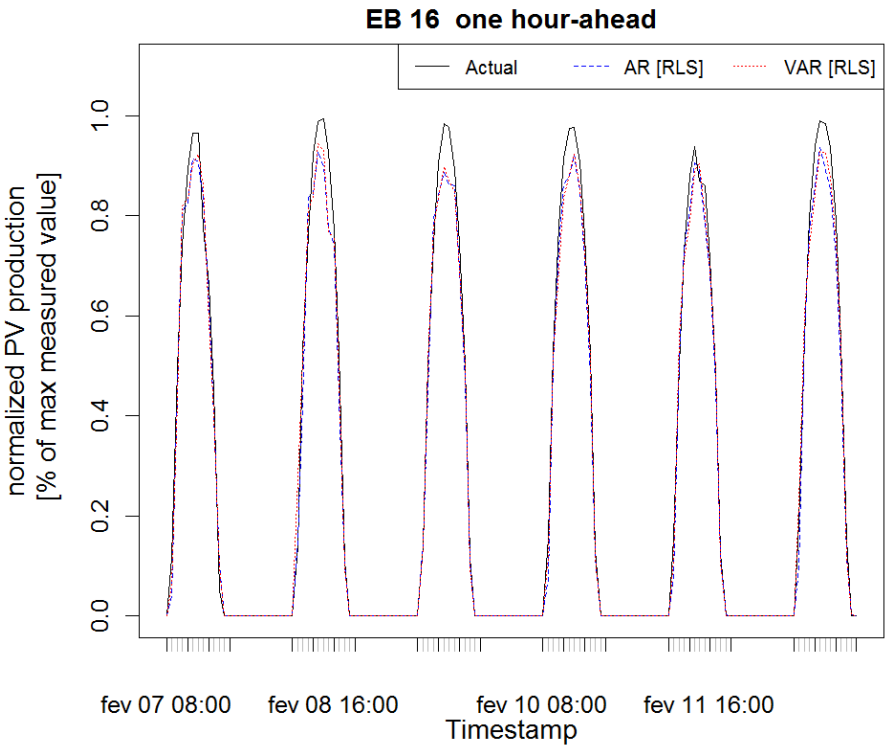


Figure 4.22 – Comparison consecutive clear sky days

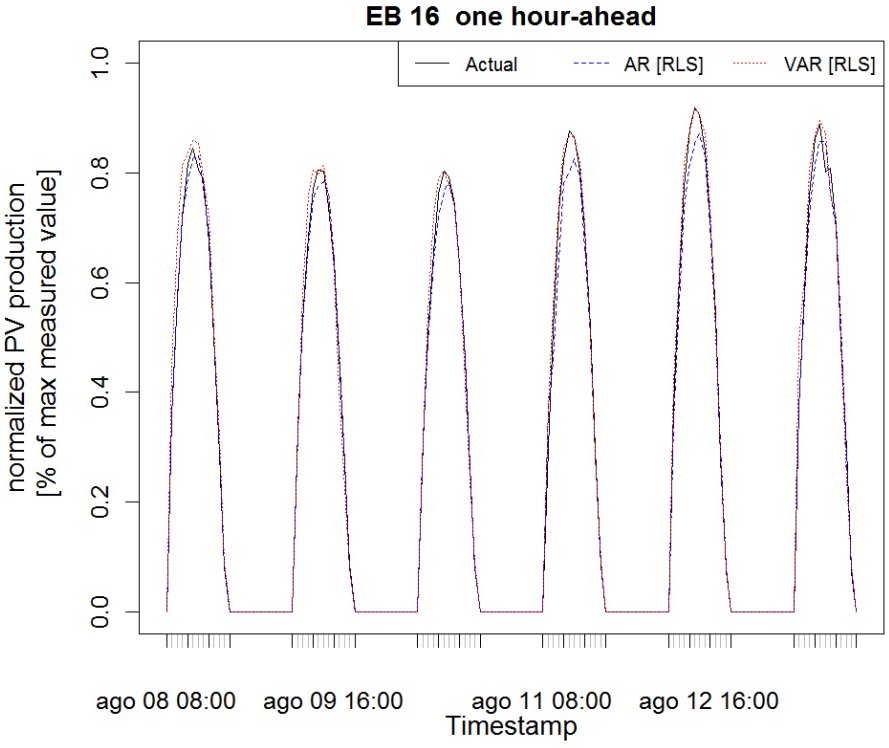


Figure 4.23 - six consecutive clear sky days

Figure 4.24 explains the difference in the forecasting quality between AR and VAR. Note that in a period with climatic instabilities, such as non-clear sky-days plus clear days with distinct cloud index, the AR predictions line shape is similar to the results obtained for the day before. Nevertheless, some magnitude adjustments seem to be done by AR model, but the

shape is clearly linked to the previous day. In this cases, the results of AR are worse than the VAR model.

Note that in Figure 4.22, Figure 4.23 and Figure 4.24, there are no zero values between consecutive days. It is explained because they have been excluded from the forecast process. In fact, only values between 7h and 19h were considered. Considering that the input variables used in AR(2) diurnal model are $p_t$, $p_{t-1}$ and $p_{(t-24+k)}$, the forecast can only be made between 9h and 19h. Thus, each "arch" represents one day.

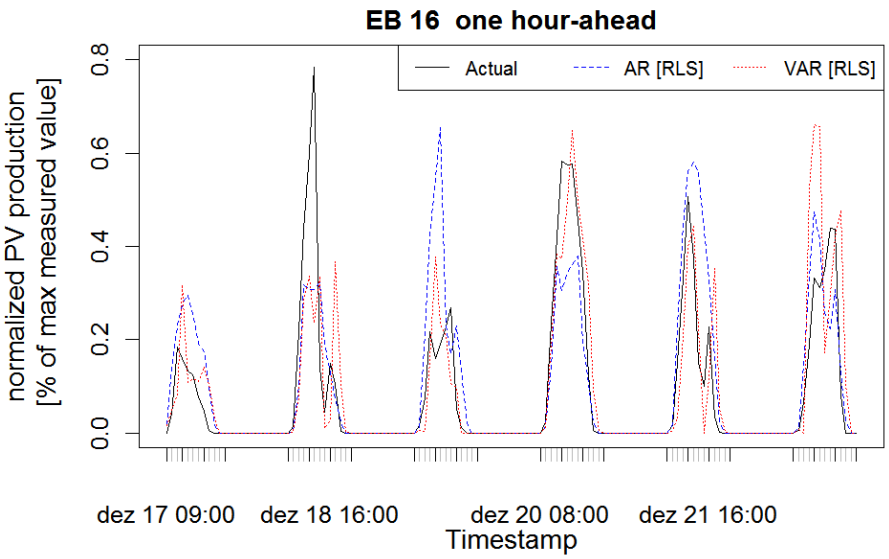


Figure 4.24 – six consecutive cloudy days

Figure 4.25 shows that DTC forecasting are more accurate and less disperse than EB predictions for adaptive models, like in Figure 4.2, and for non-adaptive models such as in Figure 4.3.

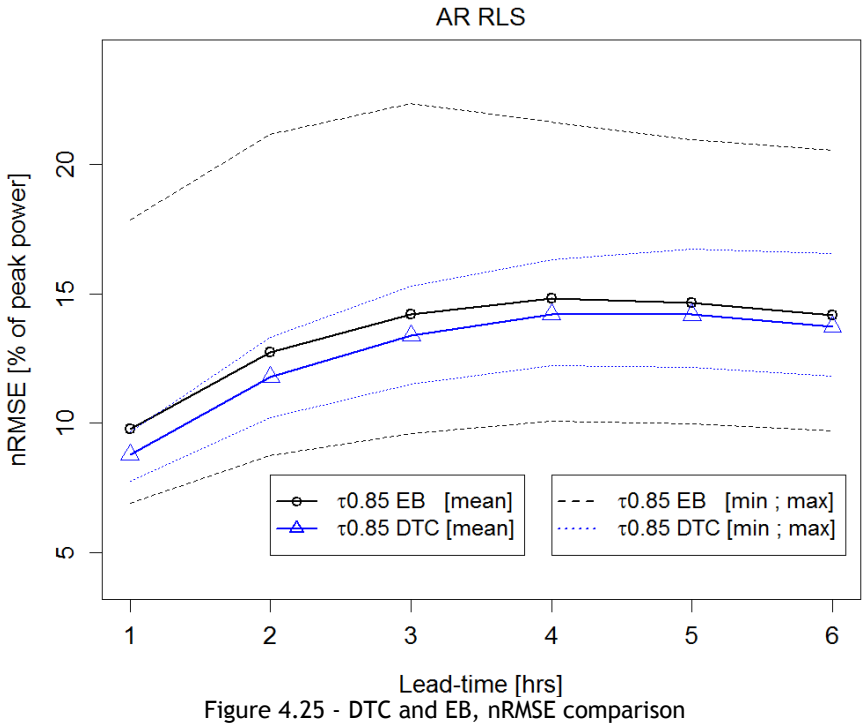


Figure 4.25 - DTC and EB, nRMSE comparison

Figure 4.26 and Figure 4.27 show that the best results are obtained using the quantile probability of 0.85,i.e., $\tau = 0.85$for online models using recursive least squares. This behavior was also obtained using off-line models at DTC level and using ordinary least squares.

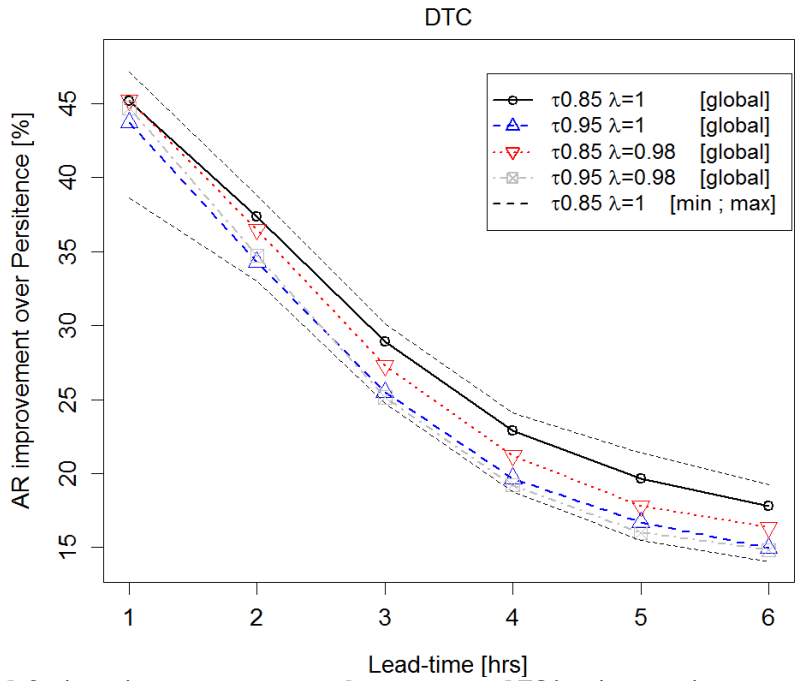


Figure 4.26 – AR(2) diurnal improvement over Persistence at DTC level, using distinct $\lambda$ and $\tau$ values.

In Figure 4.26 and Figure 4.27 the results of different $\lambda$, $\lambda = 0.98$ and $\lambda = 1$, are compared. Note that the factor $\lambda = 1$ means that every data are used in forecasting.

From Figure 4.26 and Figure 4.27, the AR(2) diurnal model for online analysis, in both DTC and EB, is the model which takes $p_t$, $p_{t-1}$ and $p_{t-24+k}$ as input and uses $\tau = 0.85$ and $\lambda = 1$.

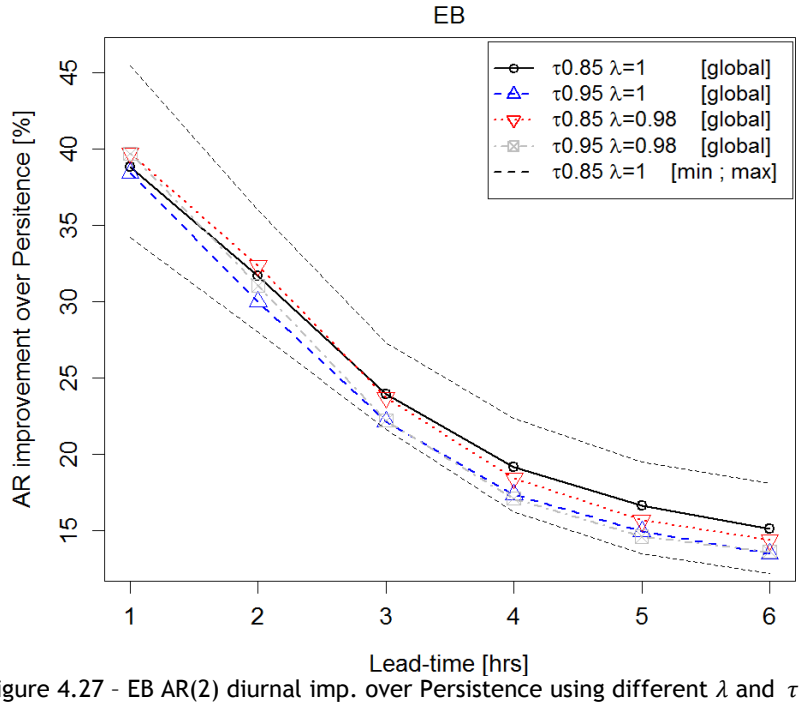


Figure 4.27 – EB AR(2) diurnal imp. over Persistence using different $\lambda$ and $\tau$

From the comparison between Figure 4.27, Figure 4.28 and Figure 4.29, one can realize the impact of the forgetting factor, $\lambda$, in the forecasting. In contrast with the EB models, the best results came from the models with $\lambda = 0.98$, except for one hour-ahead in EB 16. Note that while for $nRMSE$, the lower value is the best one, the improvement is expected to be the largest.

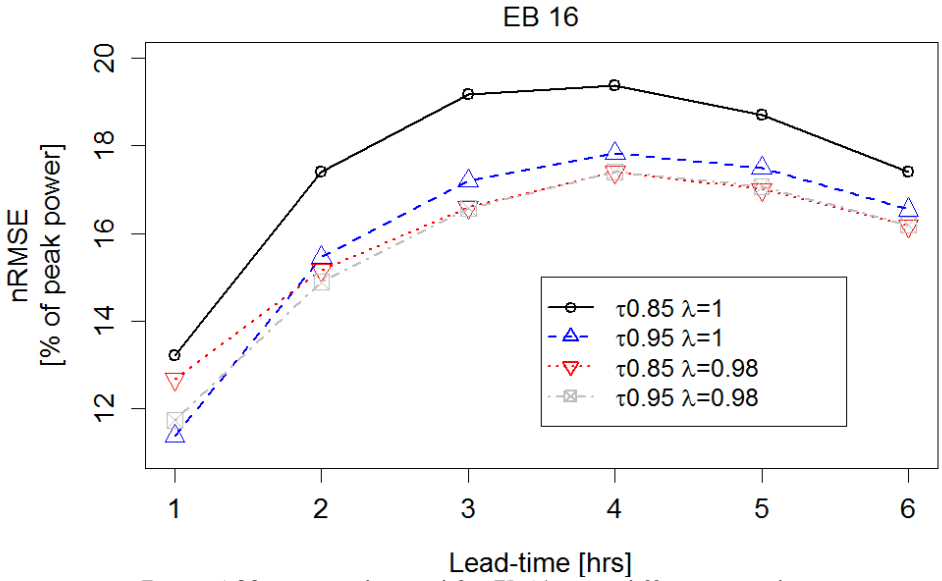


Figure 4.28 $nRMSE$ obtained for EB 16 using different $\lambda$ and $\tau$

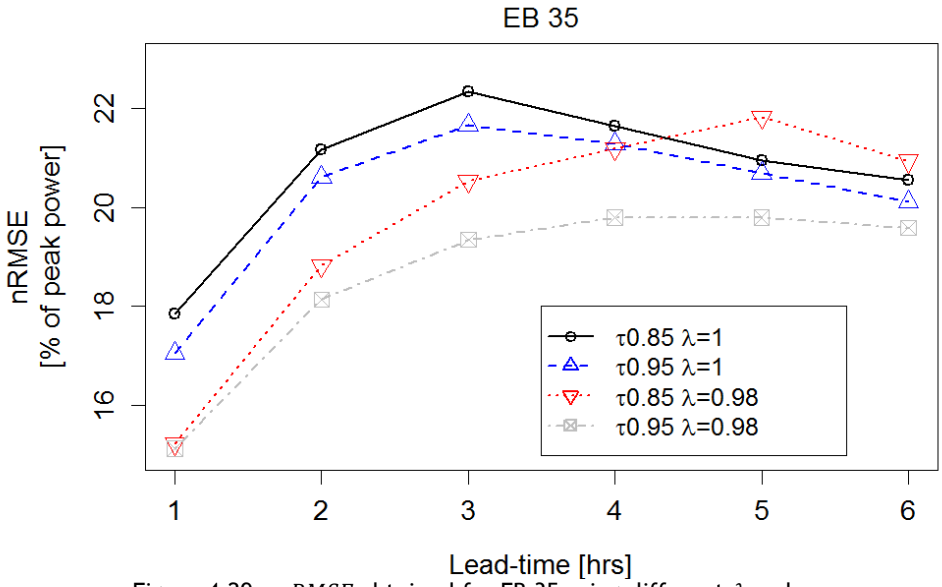


Figure 4.29 - $nRMSE$ obtained for EB 35 using different $\lambda$ and $\tau$

The gain obtained with $\lambda = 0.98$, for these two EB models, proves that $\lambda < 1$ can be useful in cases where the PV generation change with time as in EB 16 and EB 35.

Figure 4.30 shows the performance of VAR and VARX on-line models. They were tested using two different combinations of input variables. The models, tested with the AR(2) diurnal input variables, i.e., $p_t$, $p_{t-1}$ and $p_{t-24+k}$, are referred as VAR and VARX. VAR(2) and VARX(2) are second order vector autoregressive models, i.e., the seasonal term is not included as an input variable.

Figure 4.30, Figure 4.31 and Figure 4.32 show that VARX model generates a better forecasting than VAR models. Only for one-hour-ahead, the VAR(2) model presents better results than VARX(2). This behavior indicates that EB provides information that even increasing the dimension of the problem, the forecasting accuracy is improved.

The DTC number 5 shows the greatest improvement reaching 13.5% in VARX model for three hours-ahead and 5.4% in VAR model for six hours-ahead. On the other hand, DTC number 10 is the one presenting the lowest improvement with 7.1% in VARX model for three hours-ahead and 0.26% in VAR model for six hours-ahead. In this analysis, were just considered the best value obtained from VAR and VAR(2) for each time-horizon.

In global terms, the improvement using DTC models varies between 9.7% with VAR(2) model for one hour-ahead and 2.2% with VAR model for six hours-ahead. Once more, the best value between VAR and VAR(2) for each time-horizon is considered.

In contrast to the AR and VAR models, the VARX(2) model for one hour-ahead performs worse than the VARX model.

Analyzing Figure 4.30, Figure 4.31 and Figure 4.32, one can realize that as the forecasting time-horizon increases the models accuracy decreases. In particular, both VAR(2) and VARX(2) show seasonality in data series even when the clear sky model is applied.

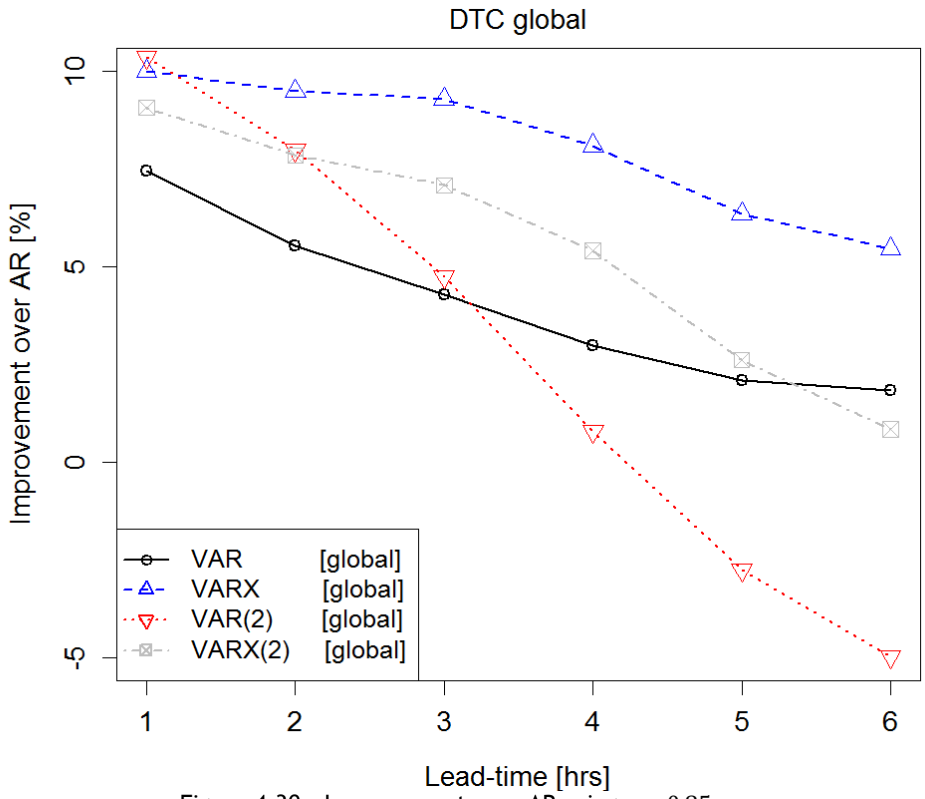


Figure 4.30 - Improvement over AR using $\tau = 0.85$.

An important information provided from Figure 4.30, Figure 4.31 and Figure 4.32 is that VAR model presents positive improvement over AR(2) diurnal model for all time horizons. The improvement values varies between 13.8%, achieved by the VARX model applied to DTC number 8 and for one hour-ahead, and 0.26% presented by VAR model applied to DTC number 10 and for five hours-ahead.

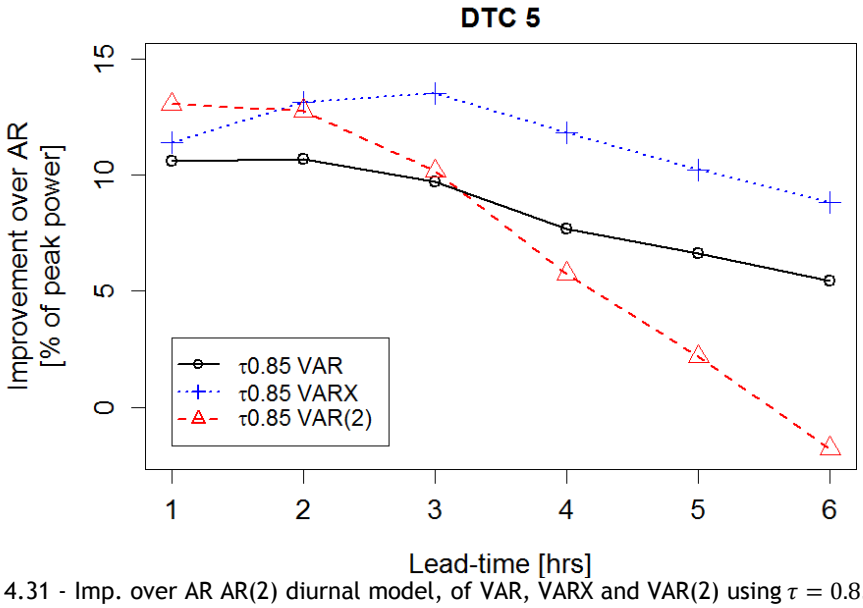


Figure 4.31 - Imp. over AR AR(2) diurnal model, of VAR, VARX and VAR(2) using $\tau = 0.85$ for one DTC.

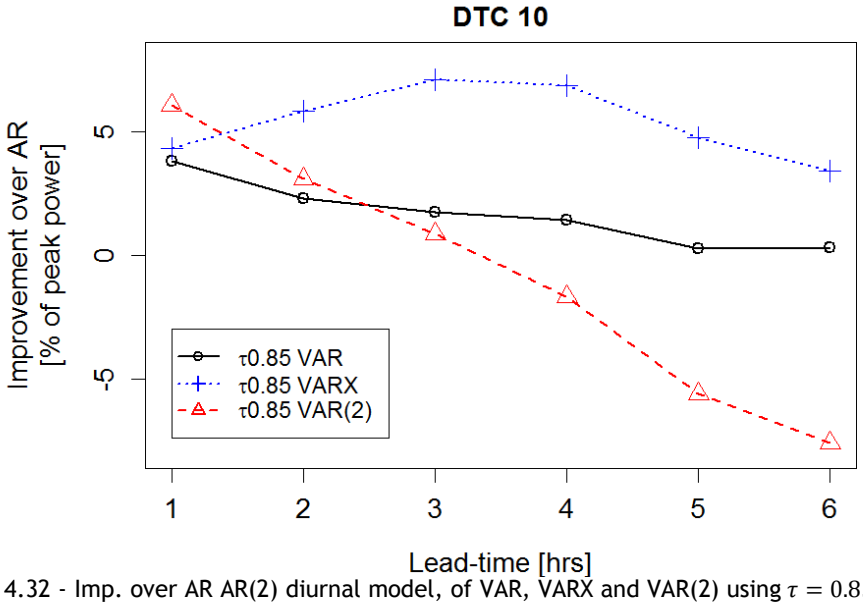


Figure 4.32 - Imp. over AR AR(2) diurnal model, of VAR, VARX and VAR(2) using $\tau = 0.85$ for one DTC.

Using different $\tau$ parameters and input variables, the Figure 4.33, Figure 4.34 and Figure 4.35.In Figure 4.33, AR, VAR and VAR(2) are compared. As the VAR model presents the best results for every time-horizon, the VAR results are exclusively presented inFigure 4.34 and Figure 4.35.

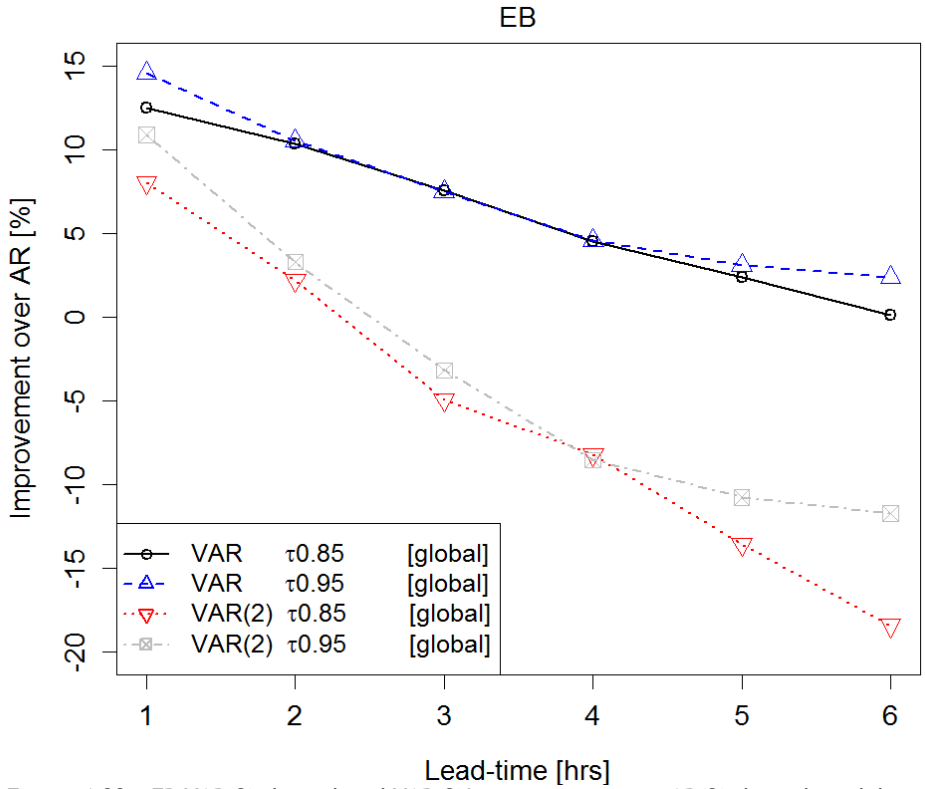


Figure 4.33 - EB VAR(2) diurnal and VAR(2)Improvement over AR(2) diurnal model.

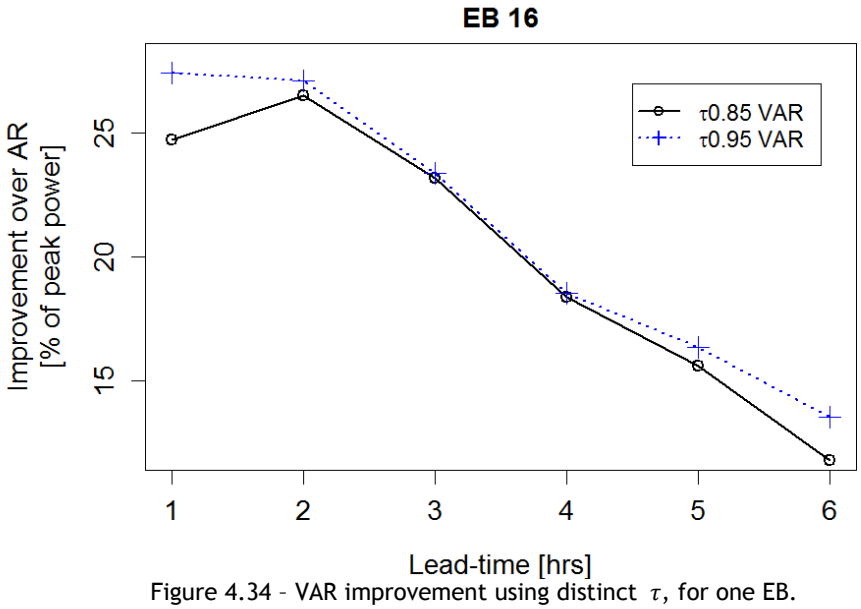


Figure 4.34 - VAR improvement using distinct $\tau$, for one EB.

Note that EB number 16, as Table 4.9 shows, has $nRMSE$ significantly higher than the average.

Although Table 4.10 shows that EB 16 presents an $nRMSE$ above the average for all time-horizon, it is closer to the average than the value obtained by the AR model. Thus, one can conclude that the VAR model is capable of use the neighbor information improving the forecasting, especially for this EB, which presented poor predictions in the AR model.

Table 4.9 - AR average $nRMSE$ and AR EB 16 $nRMSE$, $\tau = 0.85$

| | $(t+1)$ | $(t+2)$ | $(t+3)$ | $(t+4)$ | $(t+5)$ | $(t+6)$ |
|---|---|---|---|---|---|---|
| **EB 16 (%)** | 13.2 | 17.4 | 19.2 | 19.4 | 18.7 | 17.4 |
| **EB average (%)** | 9.8 | 12.8 | 14.2 | 14.8 | 14.6 | 14.2 |

Table 4.10 - VAR average $nRMSE$ and VAR EB 16 $nRMSE$, $\tau = 0.85$

| | $(t+1)$ | $(t+2)$ | $(t+3)$ | $(t+4)$ | $(t+5)$ | $(t+6)$ |
|---|---|---|---|---|---|---|
| **EB 16 (%)** | 9.9 | 12.8 | 14.7 | 15.8 | 15.8 | 15.4 |
| **EB average (%)** | 8.6 | 11.4 | 13.2 | 14.2 | 14.3 | 14.1 |

Although Figure 4.35 shows that VAR presents a positive improvement over AR, there are some cases where the improvement is negative. In fact, for tested data and using $\tau = 0.85$, 84.5% of the points achieve positive improvement assuming that each time horizon of any EB is considered as a point (44 EB and 6 time-horizons led to 264 points). $\tau = 0.95$The positive improvement percentage rise to $92.4\%$when $\tau = 0.95$ was used.

EB VAR improvement varies between $27.4\%$using EB 16 for one hour-ahead and with $\tau = 0.95$, and $-5.2\%$ using EB 44 for six hours-ahead and with $\tau = 0.85$.

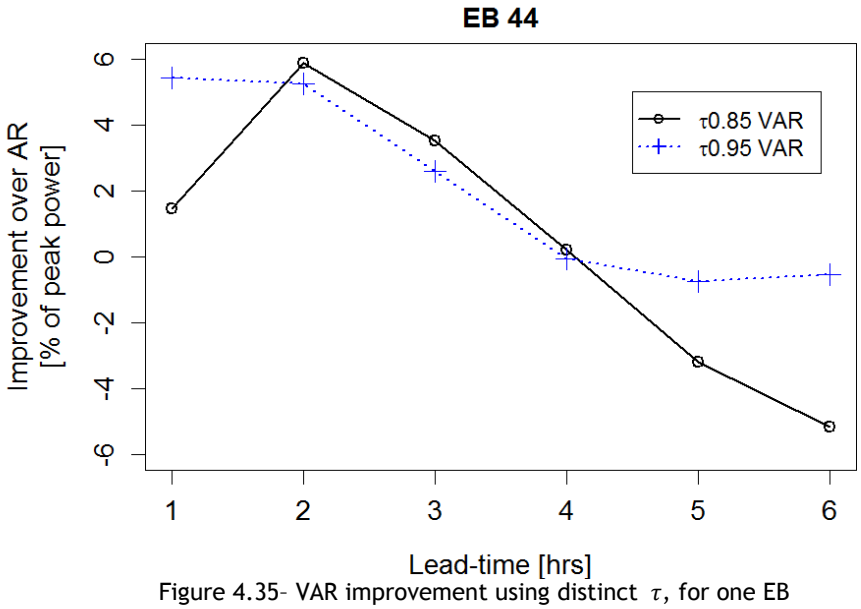


Figure 4.35- VAR improvement using distinct $\tau$, for one EB

Figure 4.36 presents the behavior of VAR model for EB and VARX model for DTC when $\tau = 0.85$was used. Note that the $nRMSE$ trend to rise when the time-horizon increases.

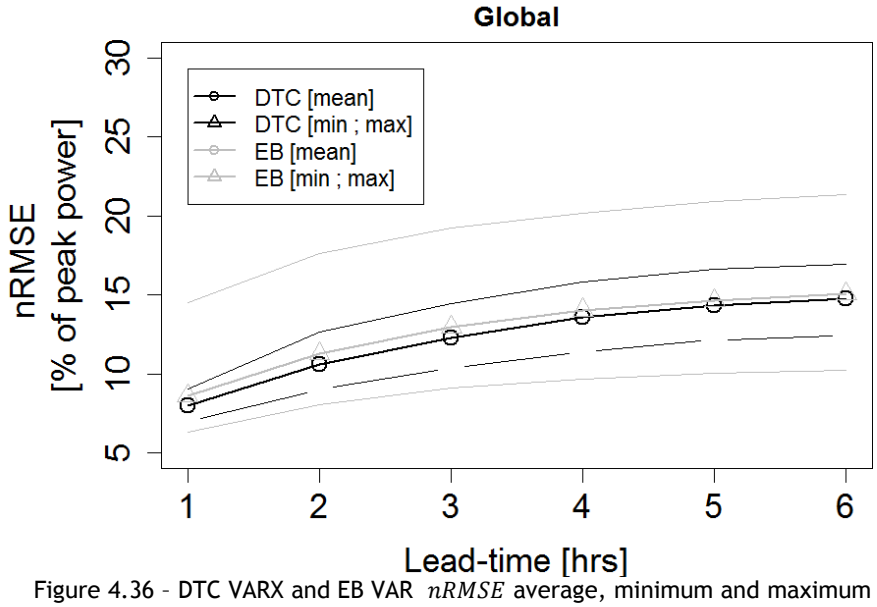


Figure 4.36 - DTC VARX and EB VAR $nRMSE$ average, minimum and maximum values

Figure 4.37 compares the behavior of the models along one year. Concerning the first two lagged hours, VAR model reaches higher improvement in winter and lower improvement in summer. Assuming that the information provided by neighborhood PV provide cloud behavior information and that winter presents more cloudy days, it is expected that neighborhood information became more important in the winter season.

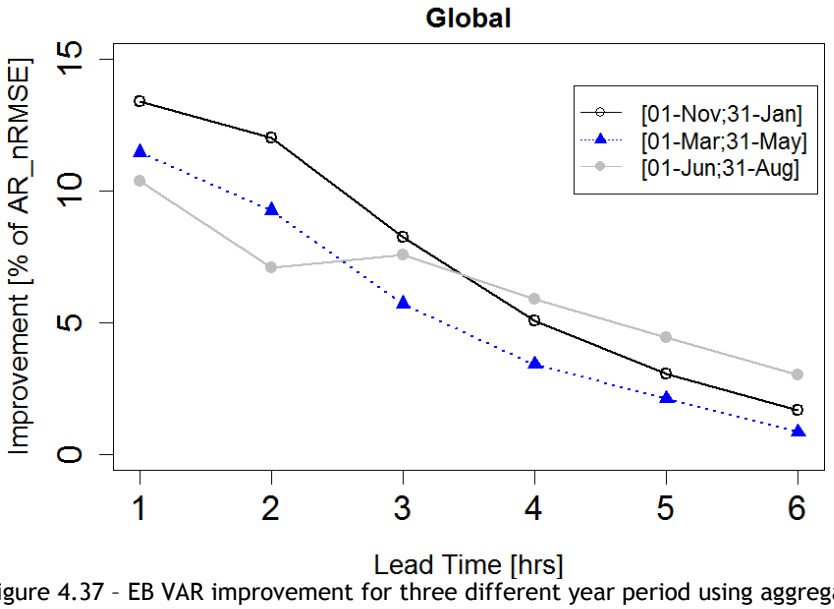


Figure 4.37 – EB VAR improvement for three different year period using aggregate EB

# Chapter 5

# Conclusions and Future Work

This chapter, firstly, section 5.1, summarizes the principal findings and conclusions drawn from the results presented in chapter 4.

Following, section 5.2 proposes a few recommendations of what could be done to continue this work.

## 5.1 Conclusions

As the Vectorial Autoregressive (VAR) models improvement over AR is positive both in on-line and off-line models, one can sentence that photovoltaic (PV) generation data of neighborhood could provide information on the impact of clouds on the generation of a region profile system improving the forecast values when compared with classical Autoregressive (AR) methods.

The results show that the introduction of the Energy Box (EB) information in the Distribution Transformer Controller (DTC) prediction brings about an improvement of about 4% in the non-adaptive, for models that do not use this information models, allowing predictions nRMSE 8.8%, whereas the adaptive models that improvement is 5.1% to 8.4% of predictions nRMSE.

The introduction of spatial and temporal information improves the overall performance by 6% in some cases enabling improvements exceeding 45% in non-adaptive models, while the adaptive models present an overall improvement to 15.4% surpassing the 25% in isolated cases.

As mentioned in previous chapter, off-line methods were tested using distinct photovoltaic normalized generation values. The production values were normalized with clear sky productions obtained using quantile probability $\tau = 0.85$ and $\tau = 0.95$. In what concerns to Gradient Boosting (GB) method beside $\tau$ parameter the methods were tested using shrinkage parameters ($\nu$) of 0.1, 0.2 and 0.5.

The results provided for GB both univariate and multivariate methods are very similar. However, GB univariate is faster, taking 8 minutes and 35 seconds to provide the same results that GB multivariate provides in 23 minutes and 35 seconds. Thus, when the results are the same GB univariate is better than GB multivariate.

Table 5.1 summarizes the off-line results at EB level. The best results, for each time horizon, are highlighted in bold. As the table presents GB performs better than OLS.

Up to two hours ahead better results are obtained with GB univariate using $\tau = 0.95$ and $\nu = 0.1$. For three and six hours-ahead better results are obtained with GB multivariate using $\tau = 0.85$ and $\nu = 0.1$. While for four and five hours-ahead better results are obtained with GB univariate using $\tau = 0.85$ and $\nu = 0.1$

Table 5.1 – EB Off-line models improvement over AR (%)

| model | | hours-ahead | one | two | three | four | five | six |
|---|---|---|---|---|---|---|---|---|
| **Off-line** | OLS | VAR $\tau = 0.95$ | 2.9 | 1.5 | 1.3 | -1.2 | -8.9 | -11.0 |
| | GB uni | VAR $\tau = 0.85\ \nu = 0.1$ | 6.9 | 6.6 | 7.7 | **6.2** | **5.1** | 3.9 |
| | | VAR $\tau = 0.95\ \nu = 0.1$ | **8.8** | **8.0** | 7.1 | 4.3 | 3.2 | 3.2 |
| | GB multi | VAR $\tau = 0.85\ \nu = 0.1$ | 5.9 | 7.3 | **8.3** | 6.2 | 5.1 | **4.5** |
| | | VAR $\tau = 0.95\ \nu = 0.5$ | 8.8 | 8.0 | 7.1 | 3.7 | 2.5 | 2.6 |

Referring to DTC level $\tau = 0.85$ performs better than $\tau = 0.95$. Table 5.2 summarizes the off-line results at DTC level. The best results, for each time horizon, are highlighted in bold. GB univariate VARX using $\nu = 0.2$ performs better for all horizons above one-hour-ahead. Up to one hour-ahead the best model is provided by GB univariate VAR using $\nu = 0.1$.

Table 5.2 – DTC Off-line models improvement over AR (%)

| model | | hours-ahead | one | two | three | four | five | six |
|---|---|---|---|---|---|---|---|---|
| **Off-line $\tau = 0.85$** | OLS | VAR | 6.3 | 4.0 | 2.8 | 2.6 | 0.0 | 2.1 |
| | | VARX | -9.5 | -1.6 | 4.2 | 5.3 | -2.7 | -2.1 |
| | GB uni | VAR $\nu = 0.1$ | **7.4** | 4.0 | 2.8 | 2.0 | 1.3 | 2.1 |
| | | VARX $\nu = 0.2$ | 6.3 | **7.9** | **7.7** | **6.6** | **5.3** | **5.5** |
| | GB multi | VAR $\nu = 0.1$ | 7.4 | 4.0 | 2.8 | 2.0 | 1.3 | 2.1 |
| | | VARX $\nu = 0.1$ | 6.3 | 7.9 | 6.3 | 4.6 | 3.3 | 3.4 |

On-line forecasting methods were tested with $\tau = 0.85$ and $\tau = 0.95$ and *forgetting factor* ($\lambda$) of 0.98 and 1.

On-line models best results were provided by the methods using $\lambda = 1$, both at EB level and DTC level.

Table 5.3 summarizes the values obtained with the Recursive Least Squares (RLS) method at EB level. The best results are provided by the VAR model using $\tau = 0.95$ and $\lambda = 1$ for all time horizons.

Table 5.3 – EB On-line models improvement over AR (%)

| model | | hours-ahead | one | two | three | four | five | six |
|---|---|---|---|---|---|---|---|---|
| On-line | RLS | VAR $\tau = 0.85$ | 12.5 | 10.4 | 7.6 | 4.5 | 2.4 | 0.1 |
| | | VAR $\tau = 0.95$ | **14.6** | **10.5** | **7.5** | **4.6** | **3.1** | **2.4** |
| | | VAR(2) $\tau = 0.85$ | 8.0 | 2.2 | -4.9 | -8.2 | -13.6 | -18.4 |
| | | VAR(2) $\tau = 0.95$ | 10.9 | 3.3 | -3.2 | -8.5 | -10.8 | -11.7 |

Table 5.4 summarizes the values obtained with the Recursive Least Squares (RLS) method at DTC level. The best results are provided using $\tau = 0.85$ and $\lambda = 1$ for all time horizons. Up to one hour-ahead the best model is the VAR(2), while for the other time horizons VAR performs better.

Table 5.4 – DTC On-line models improvement over AR (%)

| model | | hours-ahead | one | two | three | four | five | six |
|---|---|---|---|---|---|---|---|---|
| On-line | RLS | VAR $\tau = 0.85$ | 7.4 | 5.5 | 4.3 | 3.0 | 2.1 | 1.8 |
| | | VAR(2) $\tau = 0.85$ | **10.4** | 8.0 | 4.8 | 0.8 | -2.8 | -5.0 |
| | | VARX $\tau = 0.85$ | 10.0 | **9.5** | **9.3** | **8.1** | **6.4** | **5.4** |
| | | VARX(2) $\tau = 0.85$ | 9.1 | 7.8 | 7.1 | 5.4 | 2.6 | 0.8 |

## 5.2 Future work

Throughout this document has proved that GB clearly improves the forecasting in high dimensional problems. Moreover, it is also clear that on-line adaptive methodology like RLS improves the forecasting. However, some questions were not completely clear in this work. Therefore a few recommendations for future are:

- As clear sky has influence on results, the better the clear sky model, probably, the better are the results. Thus, the clear sky model should be improved. A bigger sample can be used, increasing the clear sky sample, or a more complex clear sky model capable of producing better results with the same sample;
- The literature suggests that in GB multivariate method, using as correlation matrix the covariance matrix instead identity matrix provide better results. In tests made

throughout this work, identity matrix provides lower forecasting errors, it suggests that different methods for estimating the covariance matrix can be tested;

- Models accuracy reduces for longer time horizons. The introduction of Numerical Weather Predictions (NWP) as models input, could improve the forecasting accuracy for longer time horizons.
- Develop an adaptive on-line model based on GB method will probably allow to reach better forecasting. Therefore, also suggests the development of an adaptive method based on GB;
- All predictions presented in this dissertation are deterministic. In the forecasting industry, it is currently used probabilistic forecasting, to include in decision-making models, for instance. Thus suggests the development of a probabilistic forecasting model.

# Bibliography


[1] G. Masson, M. Latour, M. Rekinger, I.-T. Theologitis, and M. Papoutsi, “Global Market Outlook.” 2013.

[2] P. L. Monteiro, “Projeto InovGrid,” *Renováveis Magazine*, pp. 50–53, 2012.

[3] C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, and G. Conzelmann, “Wind power forecasting: state-of-the-art 2009.” 2009.

[4] R. Neves, “Desenvolvimento de Modelos de Previsão de Produção de Centrais Solares Fotovoltaicas,” FEUP, 2010.

[5] L. Fernandez-Jimenez, A. Muñoz-Jimenez, A. Falces, M. Mendoza-Villena, E. Garcia-Garrido, P. M. Lara-Santillan, E. Zorzano-Alba, and P. J. Zorzano-Santamaria, “Short-term power forecasting system for photovoltaic plants,” *Renew. Energy*, vol. 44, pp. 311–317, Aug. 2012.

[6] E. Lorenz, J. Remund, S. C. Müller, W. Traunmüller, G. Steinmaurer, D. Pozo, J. Antonio, V. L. Fanego, L. Ramirez, M. G. Romeo, C. Kurz, L. M. Pomares, and C. G. Guerrero, “Benchmarking of different approaches to forecast solar irradiance.” .

[7] J. Remund, R. Perez, and E. Lorenz, “Comparison of solar radiation forecasts for the USA,” *Eur. PV Conf.*, vol. 2, pp. 3–5, 2008.

[8] R. Perez, M. Beauharnois, Karl Hemker Jr., S. Kivalov, E. Lorenz, S. Pelland, J. Schlemmer, and G. Van Knowe, “Evaluation of numerical weather prediction solar irradiance forecasts in the US,” in *American Solar Energy Society - Proc. ASES Annual Conference*, 2011.

[9] R. Zamora, E. Dutton, M. Mckeen, J. Wilczak, and Y.-T. Hou, “The accuracy of solar irradiance calculations used in mesoscale numerical weather prediction,” *Mon. Weather Rev.*, vol. 133, pp. 783–792, 2005.

[10] P. Bacher, H. Madsen, B. Perers, and H. A. Nielsen, “A non-parametric method for correction of global radiation observations,” *Sol. Energy*, vol. 88, pp. 13–22, Feb. 2013.

[11] H. Ohtake, K. Shimose, J. G. D. S. Fonseca, T. Takashima, T. Oozeki, and Y. Yamada, “Accuracy of the solar irradiance forecasts of the Japan Meteorological Agency mesoscale model for the Kanto region, Japan,” *Sol. Energy*, vol. 98, pp. 138–152, Dec. 2013.

[12] M. Wittmann, H. Breitkreuz, M. Schroedter-Homscheidt, and M. Eck, “Case Studies on the Use of Solar Irradiance Forecast for Optimized Operation Strategies of Solar Thermal Power Plants,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 1, no. 1, pp. 18–27, Mar. 2008.

[13] A. Hammer, D. Heinemann, C. Hoyer, R. Kuhlemann, E. Lorenz, R. Müller, and H. G. Beyer, “Solar energy assessment using remote sensing technologies,” *Remote Sens. Enviorment*, vol. 86, no. 3, pp. 423–432, Aug. 2003.

[14] M. Ahlstrom and J. Kankiewicz, “Perspective and understanding on solar power forecasting,” in *Solar Power Forecasting*, 2009.

[15] H. Diagne, M. David, P. Lauret, and J. Boland, “Solar Irradiation Forecasting: State-of-the-art and Proposition for Future Developments for Small-scale Insular Grids,” *reuniwatt.com*, pp. 1–8, 2012.

[16] P. Bacher, H. Madsen, and H. Nielsen, “Online short-term solar power forecasting,” *Sol. Energy*, vol. 83, no. 10, pp. 1772–1783, Oct. 2009.

[17] H. Pedro and C. Coimbra, “Assessment of forecasting techniques for solar power production with no exogenous inputs,” *Sol. Energy*, vol. 86, no. 7, pp. 2017–2028, Jul. 2012.

[18] A. Mellit and A. M. Pavan, “A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy,” *Sol. Energy*, vol. 84, no. 5, pp. 807–821, May 2010.

[19] J. Huang, M. Korolkiewicz, M. Agrawal, and J. Boland, “Forecasting solar radiation on an hourly time scale using a Coupled AutoRegressive and Dynamical System (CARDS) model,” *Sol. Energy*, vol. 87, pp. 136–149, Jan. 2013.

[20] W. Ji and K. C. Chee, “Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN,” *Sol. Energy*, vol. 85, no. 5, pp. 808–817, May 2011.

[21] R. Marquez and C. F. M. Coimbra, “Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database,” *Sol. Energy*, vol. 85, no. 5, pp. 746–756, May 2011.

[22] C. Paoli, C. Voyant, M. Muselli, and M.-L. Nivet, “Forecasting of preprocessed daily solar radiation time series using neural networks,” *Sol. Energy*, vol. 84, no. 12, pp. 2146–2160, Dec. 2010.

[23] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, “Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 2, no. 1, pp. 2–10, Mar. 2009.

[24] E. Lorenz, T. Scheidsteger, J. Hurka, D. Heinemann, and C. Kurz, “Regional PV power prediction for improved grid integration,” *Prog. Photovoltaics Res. Appl.*, vol. 19, no. September 2010, pp. 757–771, 2011.

[25] C. Chen, S. Duan, T. Cai, and B. Liu, “Online 24-h solar power forecasting based on weather type classification using artificial neural network,” *Sol. Energy*, vol. 85, no. 11, pp. 2856–2870, Nov. 2011.

[26] A. Sfetsos and A. H. Coonick, “Univariate and Multivariate Forecasting of Hourly Solar Radiation with Artificial Intelligence Techniques,” *Sol. Energy*, vol. 68, no. 2, pp. 169–178, 2000.

[27] C. Silva, “Desenvolvimento de uma metodologia e ferramentas para a previsão da produção elétrica em parques fotovoltaicos,” FEUP, 2012.

[28] V. Berdugo, C. Chaussin, L. Dubus, G. Hebrail, and V. Leboucher, “Analog Method for Collaborative Very-Short-Term Forecasting of Power Generation from Photovoltaic Systems,” *kd2u.org*.

[29] C. Yang and L. Xie, “A novel ARX-based multi-scale spatio-temporal solar power forecast model,” in *North American Power Symposium (NAPS), 2012*, 2012.

[30] H. Von Storch and F. Zwiers, *Statistical analysis in climate research*. CAMBRIDGE UNIVERSITY PRESS, 2001.

[31] P. Bacher, “Short-term solar power forecasting,” 2008.

[32] D. N. . Gujarati and D. C. Porter, *Basic Econometrics*, Fifth inte. New York: McGraw-Hill, 2009.

[33] A. Zellner, “An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias,” *J. Am. Stat. Assoc.*, vol. 57, no. 298, pp. 348–368, 1962.

[34] L. Ljung and T. Soderstrom, *Theory and practice of recursive identification*. London, England: The MIT Press, Cambridge, Massachusetts, 1983.

[35] R. Lutz and P. Buhlmann, “Boosting for high-multivariate responses in high-dimensional linear regression,” *Stat. Sin.*, vol. 16, pp. 471–494, 2006.

[36] J. Friedman, “Greedy function approximation: a gradient boosting machine,” *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001.

[37] P. Bühlmann, “Boosting for high-dimensional linear models,” *Ann. Stat.*, vol. 34, no. 2, pp. 559–583, Apr. 2006.

# Appendix

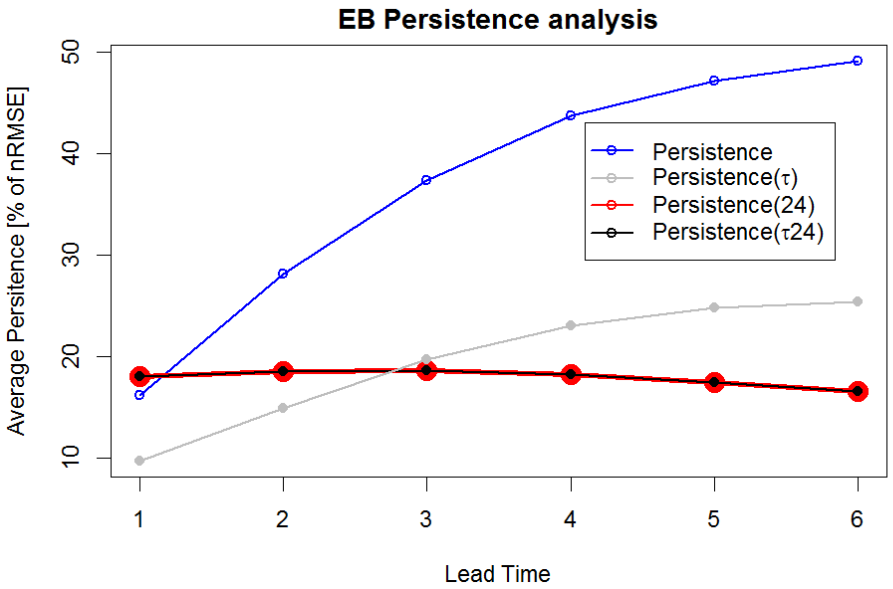


Figure A.1 - Comparison of nRMSE obtained up to six hours ahead for EB

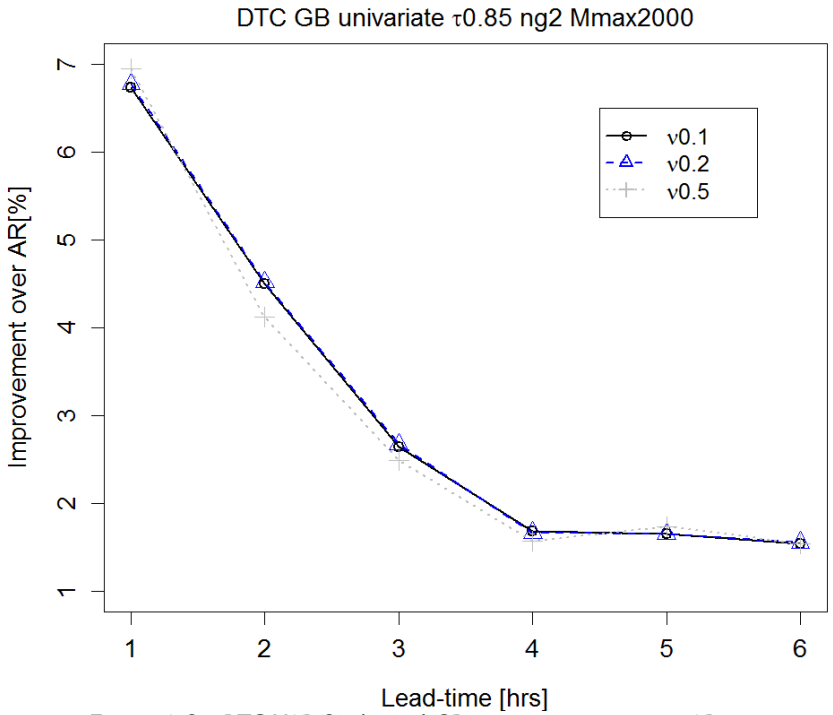


Figure A.2 – DTC VAR(2) diurnal GB univariate imp over AR

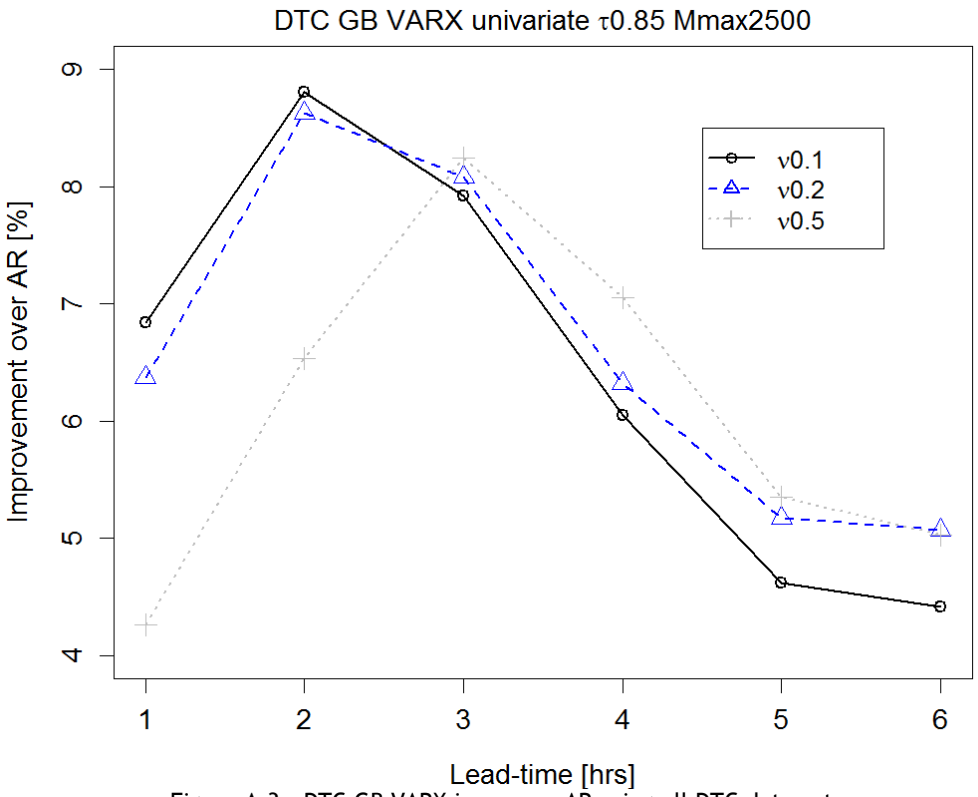


Figure A.3 – DTC GB VARX imp. over AR using all DTC data set

# Solar Power Forecasting in Smart Grids Using Distributed Information

R.J. Bessa, A. Trindade, A. Monteiro, V. Miranda
INESC TEC - INESC Technology and Science and FEUP - Faculty of Engineering, University of Porto
Porto, Portugal
rbessa@inescporto.pt, ee08089@fe.up.pt,
andre.b.monteiro@inescporto.pt, vmiranda@inescporto.pt

Catia S. P. Silva
NeuroEngineering Laboratory, University of Florida
Gainesville, FL USA
catiaspsilva@ufl.edu

***Abstract*— The growing penetration of solar power technology at low voltage (LV) level introduces new challenges in the distribution grid operation. Across the world, Distribution System Operators (DSO) are implementing the Smart Grid concept and one key function, in this new paradigm, is the solar power forecasting. This paper presents a new forecasting framework, based on the vector autoregression theory, that combines spatial-temporal data collected by smart meters and distribution transformer controllers to produce six-hour-ahead forecasts at the residential solar photovoltaic (PV) and secondary substation (i.e., MV/LV substation) levels. This framework has been tested for 44 micro-generation units and 10 secondary substations from the Smart Grid pilot in Évora, Portugal (one demonstration site of the EU Project SuSTAINABLE). A comparison was made with the well-known Autoregressive forecasting Model (AR – univariate model) leading to an improvement from 3% to 25% for a six hour time-horizon.**

***Keywords—Solar power; forecasting; smart grid; spatial-temporal; smart metering***

## I. INTRODUCTION

Several countries are achieving remarkable levels of installed solar power. By the end of 2012, the installed solar power in Germany and Italy was around 32.6 GW and 16.7 GW respectively. In terms of cost, solar photovoltaic (PV) is reaching grid parity in many countries, meaning that it can generate electricity at a levelized cost less than or equal to the electricity retailing tariffs [1]. In this context, the deployment of solar PV will likely continue even if subsidies are withdrawn (e.g., feed-in tariffs).

The majority of this installed capacity is connected to the medium and low voltage (MV and LV) distribution grids. The roll-out of the Smart Grid concept [2], strongly supported by Information and Communication Technologies (ICT), provides additional capabilities for monitoring and controlling the distribution grid and its assets. This creates conditions to develop new management tools (or improve existing ones) that maximize the integration of distributed generation at the MV and LV levels [3], such as voltage control [4], state-estimation [5] and congestion management [6].

Furthermore, a massive deployment of local storage (such as batteries) at the residential level might occur if governments create incentives for such goal [7]. For instance, Germany has created financial incentives for owners of solar systems with batteries [8].

The abovementioned management tools and the joint coordination of PV panels and local storage require the use of solar power forecasts for several hours ahead. The time-horizon of interest for power system operations can be divided into two classes [9]: (a) very short-term (up to six hours ahead); (b) short-term (up to three days ahead).

In recent publications, several works combine statistical algorithms with numerical weather predictions (NWP) to produce solar power forecasts for the short-term horizon. Bacher *et al.* [10] describe a two-stage forecasting approach. First, a clear sky model, based on weighted quantile regression, is proposed to remove the diurnal component of solar generation and global irradiance. Then, an autoregressive model with exogenous inputs (ARX) is used to combine past observations of solar power with NWP. According to the authors, up to two hours ahead the most important inputs are the past observations, while for a horizon up to 36 hours ahead the NWP prevail. Fernandez-Jimenez *et al.* [11] also used NWP as input in several machine learning algorithms (i.e., Auto-Regressive Integrated Moving Average – ARIMA, k-nearest neighbors - kNNs, neural networks – NN, and adaptive neuro-fuzzy models) to produce solar power forecasts for the next 39 hours. The best performance was obtained with a Neural Network (NN).

For the very short term horizon, two different classes of models can be found. The first class is based on satellite images. Hammer *et al.* [12] describe an algorithm based on cloud-index images that are predicted with motion vector fields derived from two consecutive images.

This work was made in the framework of the BEST CASE project ("NORTE-07-0124-FEDER-000056") financed by the North Portugal Regional Operational Programme (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF), and by national funds, through Fundação para a Ciência e a Tecnologia (FCT). It was also co-financed by the 7th RTD Framework Programme within the SuSTAINABLE project (contract no. 308755), and by the COMPETE Programme and the FCT within projects «SMAGIS – PTDC/SEN-ENR/113094/2009» and «DYMONDS – CMU-PT/SIA/0043/2009».

The second class consists of univariate time-series models. Pedro and Coimbra [13] compared the performance of different machine learning algorithms (i.e., ARIMA, kNNs, NN and NN optimized by genetic algorithms), which only use past observations of the time-series as inputs. The NN, combined with genetic algorithms, obtained the best performance.

An extended overview for the solar power forecast literature can be found in [14].

The forecasting framework presented in this paper addresses the very short-term horizon and is included in the second class of models. The main difference is that the new proposed method is spatial-temporal, since it combines the past observations of time series distributed in space. To our knowledge, only two works combined information from neighboring sites to improve solar power forecast. Berdugo *et al.* [15] described a method based on searching similar local and global current states (e.g. kNNs), considering neighbor sites, whose goal is not to produce the "optimal" forecast (i.e., with minimum error); instead, it handles distributed data streams and maintains power measurements' privacy. Yang *et al.* [16] proposed an ARX model for each solar site where the exogenous variables are measurements from neighbor sites.

This paper proposes three main contributions:

a) new forecasting method, constructed on the top of a Smart Grid infrastructure, based on vector auto-regression theory (VAR), combining information from the distributed PV panels;

b) improved accuracy by introducing exogenous variables to the model, i.e. observations from micro-generation smart meters;

c) an online fitting method, based on recursive least squares, for the VAR model.

This forecasting framework will be applied to produce six-hour-ahead forecasts for each residential PV and secondary substation (i.e., MV/LV). The proposed method can operate as a centralized forecasting system to be used by a Distribution System Operator (DSO) for managing distributed energy resources or by a solar power aggregator for participating in intra-day electricity markets.

The paper is organized as follows: section 2 presents the Smart Grid infrastructure in Portugal; section 3 describes the solar power forecasting algorithms; the test case results are presented in section 4; section 5 presents the conclusions and future work.

## II. SMART GRID INFRASTRUCTURE

During the recent years, several DSO have conducted different Smart Grid pilot tests and a deployment of these technological solutions is expected for the following years. An example is the InovGrid Project in Portugal [17], promoted by EDP Distribution, aiming to develop new ICT technology and computational tools for automating network management and to create a complete smart distribution grid. This project resulted in a large-scale demonstration pilot in the city of Évora in Portugal, named InovCity [18], which is also one demonstration site of the EU Project SuSTAINABLE.

The main components of this infrastructure are the EDP Box (EB) and the Distribution Transformer Controller (DTC). The EB is a smart meter with load and generation management functions, located at each delivery point. Load and generation are metered separately. It can interact with other devices through a home area network. The DTC is located at the secondary substation level, comprising modules for measurement, remote control and communication actions. It collects data from the EB and the secondary substation. Both the EB and the DTC are part of a hierarchical control and communication architecture. Each EB has a bi-directional communication with the corresponding DTC through GPRS (General Packet Radio Service) or PLC (Power Line Communications), and the DTC communicates with the SCADA/DMS through a wide area network based on GPRS [18]. Other communication technologies, such as radio frequency mesh modules, can be explored in this framework.

At the primary substation level (HV/MV), a Smart Substation Controller (SSC) is installed. The SSC is responsible for aggregating and managing the operational data from EB and DTC, and for applying demand-side management, self-healing and generation management strategies. Therefore, the MV grid is managed by the SSC. On the top of these technologies, there are services capable of handling large volumes of data and, at the same time, providing an overview of all existing devices.

Section 3 presents the forecasting framework that follows the Smart Grid infrastructure of the InovCity pilot. The forecasting system is installed at the central management level (i.e., in the DMS). Although installed in the DMS, the forecasting system can be virtually distributed by HV/MV substation, but using information from DTC connected to different primary substations.

The outputs are forecasts for each DTC and EB. For the forecast at the DTC level, the EB measurements can be used as distributed sensors to better capture the effect of clouds in solar generation and consequently improve the forecasts, which in turn increase the amount of transmitted data. Section 4 evaluates the improvement that is obtained with this additional information.

Finally, it is important to stress that the forecasting system is also valid for a non-hierarchical architecture, such as one where each device communicates directly with the DSO control center via TCP/IP [19]. One has to pose a centralized data flow topology as a requirement. A forecasting approach for peer-to-peer data flow is presented in [14].

## III. FORECASTING FRAMEWORK

### A. Seasonal De-Trending of the Time-Series

The solar power time series present a seasonal pattern dependent on the time of the day and day of the year, which results in a trend component during the day. In the literature, different physical and statistical approaches are proposed to

estimate the deterministic variation of the solar irradiance (i.e., excluding the influence of clouds and other factors) [14].

One of these statistical approaches is the clear-sky model described in [10] that, based on weighted quantile regression, is directly applied to solar power time-series. The method is described as a statistical normalization, capable of generating a stationary time-series with normalized solar power. Note that most classical models, such as AR and VAR, assume stationarity of the time-series.

Following the results in [10] and [20], the predictors of the clear-sky model are the time of the day (h) and day of the year (doy). The clear-sky generation ($\hat{P}_t^{cs}$) is estimated as a local constant model and the weighted quantile regression for quantile $\tau$ can be expressed as:

$$\hat{P}_t^{cs} = \arg\min_{\hat{P}_t^{cs}} \sum_{i=1}^{N} K(h_t, doy_t, h_i, doy_i) \cdot \rho(\tau, e_i) \quad (1)$$

with $e_i = P_t - \hat{P}_t^{cs}$, where

$$K(h_t, doy_t, h_i, doy_i) = \frac{K(h_t, h_i, \sigma_h) \cdot K(doy_t, doy_i, \sigma_{doy})}{\sum_{i=1}^{N} \left[ K(h_t, h_i, \sigma_h) \cdot K(doy_t, doy_i, \sigma_{doy}) \right]} \quad (2)$$

is the kernel product of the two predictors which locally weights each observation, and

$$\rho(\tau, e_i) = \begin{cases} \tau \cdot e_i & , e_i \geq 0 \\ (1-\tau) \cdot e_i & , e_i < 0 \end{cases} \quad (3)$$

is the loss function of the quantile regression problem [21]. Since both variables are circular, a circular kernel is used:

$$K(x_t, x_i, \sigma) = e^{\frac{1}{\sigma} \cdot \cos\left[ 2\pi \cdot \frac{(x_t - x_i)}{d} \right]} \quad (4)$$

where $\sigma$ is the smoothing parameter and $d$ is the period of variable $x$ (e.g., equal to 24 in the time of the day).

The output of the model from Eq. 1 is used to normalize the measured solar power ($P_t$) as follows:

$$P_t^{norm} = \frac{P_t}{\hat{P}_t^{cs}} \quad (5)$$

The model's parameters are the kernel bandwidths $\sigma_h$ and $\sigma_{doy}$, as well as the quantile $\tau$. These parameters are determined by trial-error experiences, inducing a value equal to 1 for the variable $P_t^{norm}$ during clear-sky days.

### *B. Vector Autoregressive (VAR)Model*

The ARIMA process is a well-known class for univariate time-series models [22]. A special model of this class is the autoregressive (AR), in which the value of the response variable for time interval $t$ can be interpreted as a regression on past observations (or lags) of the time-series. For one hour-ahead forecast, the AR model is:

$$P_{t+1} = \alpha + \beta_1 \cdot P_t + \beta_2 \cdot P_{t-1} + \cdots + \beta_l \cdot P_{t-l} + e_{t+1} \quad (6)$$

where $\beta$ are the model's coefficients, $\alpha$ a constant term, $l$ the order of the AR model and $e_{t+1}$ is a white noise process with zero mean and constant variance $\sigma_e^2$.

The main limitation of this model is that it only uses, as predictors, the past observations from the dependent variable. This model can be extended with exogenous variables (such as NWP), forming an ARX model. Nevertheless, as mentioned in the literature [10] and [14], NWP can only improve the forecast error for time horizons greater than four hours-ahead, while for shorter time horizons the relevant information consists in time-series observations.

In order to improve the forecasts for the very short-term horizon, a VAR model [23] is used to combine past observations from the solar power in each site with past values from neighbor sites. This consists in a multi-output (or spatial-temporal) linear regression model with $N$ observations, $q$-dimensional response and $p$-dimensional predictors.

In matrix form, for one step-ahead forecast, it is given by:

$$P_{t+1} = \alpha + \mathrm{B} \cdot P_{t-l} + \mathrm{E}_{t+1} \quad (7)$$

where $P_{t+1}$ is the response matrix with dimension $N.q$, $B$ the coefficient matrix with dimension $p.q$, $P_{t-l}$ the predictor matrix with dimension $N.p$, $\alpha$ is a vector with $q$ intercept (or constant) terms, $E_{t+1}$ is a matrix with dimension $N.q$ containing i.i.d. residuals with zero mean and constant covariance $\Sigma_e$.

Eq. 7, for an example with two response variables and two lagged terms, becomes:

$$\begin{aligned} P_{t+1,1} &= \alpha_1 + \beta_{11} \cdot P_{t,1} + \beta_{12} \cdot P_{t-1,1} + \beta_{13} \cdot P_{t,2} + \beta_{14} \cdot P_{t-1,2} + e_{t,1} \\ P_{t+1,2} &= \alpha_2 + \beta_{21} \cdot P_{t,1} + \beta_{22} \cdot P_{t-1,1} + \beta_{23} \cdot P_{t,2} + \beta_{24} \cdot P_{t-1,2} + e_{t,2} \end{aligned} \quad (8)$$

As shown in Eq. 8, the VAR consists of linear univariate regression models, in which the $P_{t+1}$ of each site depends on a constant term and lagged terms of the $q$ response variables. Note that each regression equation takes the same matrix of predictors ($P_{t-l}$) and a vector of coefficients (a row from matrix $B$) is estimated from data. However, this model is capable of modeling the dynamic relation among spatially distributed time-series. In fact, Eq. 7 and 8 have the form of a Seemingly Unrelated Regression (SUR) model [23].

Estimating the coefficients of the VAR model is straightforward, that is, the Ordinary Least Squares (OLS) can be applied if the same predictors appear in every equation. In such a case, OLS gives the same solution of Generalized Least Squares (GLS) and can be applied independently to each regression equation.

### *C. Recursive Least Squares*

The ICT infrastructure of a Smart Grid generates a continuous stream of data that must be handled in quasi real-time and with low data storage requirements. The AR and VAR models described in the previous section can be fitted in quasi real-time using the recursive least squares (RLS) method with a forgetting factor [24]. This method overcomes the

problem of handling "big data" since it is not necessary to store historical data for fitting (or re-fitting) the model. Furthermore, the RLS method, with a *forgetting factor* $\lambda$, tracks changes in the dynamics of the data generating structure, such as loss of performance due to dust in PV panels or changes in the surrounding environment (e.g., shadows).

Since both VAR and AR can be fitted with OLS, the RLS method can also be applied to this model and it is of great importance since the spatial-temporal relation between PV sites is very dynamic and requires time-varying coefficients.

The update of the parameters is performed with the RLS method as follows:

$$\beta_t = \beta_{t-1} + K_t \cdot \left[ P_t - \left( \alpha + \sum_{j=1}^{l} \beta_{t-1} \cdot P_{t-l} \right) \right] \quad (9)$$

where $K_t$ is given by,

$$K_t = Q_t \cdot P_{t-l} \quad (10)$$

and $Q_t$ by

$$Q_t = \frac{1}{\lambda} \cdot \left[ Q_{t-1} - \frac{Q_{t-1} \cdot P_{t-l} \cdot P_{t-l}^T \cdot Q_{t-1}}{\lambda + P_{t-l}^T \cdot Q_{t-1} \cdot P_{t-l}} \right] \quad (11)$$

A *forgetting factor* equal to 1 leads to a recursive estimation of the coefficients, while a smaller value discounts old data with an exponential decay.

This fitting method perfectly copes with streaming data since at time step $t$ only $\beta_t$, $K_t$ and $Q_t$ have to be stored in memory, and the remaining data is dropped.

This algorithm requires some initial values for $\beta_0$ and $Q_0$. A simple and robust approach is to initialize $\beta_0$ with zeros and $Q_0$ as a diagonal matrix with a large constant value.

### *D. VAR Model for Solar Power Forecast*

In order to apply the forecasting techniques from Eq. 6 and 7, in a first phase, it is necessary to normalize the solar power time-series with the clear-sky model from section III.A. In [20], it is recommended to remove the small $\hat{P}_t^{cs}$ values, since for these values $P_t^{norm}$ increases considerably and reaches infinity in the night time.

In this paper, a different approach that enables the use of the VAR model was followed: the normalized solar power values outside the period between 7h00 and 19h00 (i.e., the average period with almost no solar generation during the whole year in Portugal) are removed from all sites.

In a second phase, the normalized solar power values are used to fit the AR and VAR models with RLS. Both models are applied to forecast the solar power for each DTC and EB. Furthermore, when specifying an AR or VAR model, it is important to determine how many lagged terms need to be included. This can be achieved with the following method: first, the autocorrelation plot of the normalized time series is analyzed to make a coarse estimation of the necessary lags; then, the autocorrelation plot of the residuals is analyzed to check if the residuals are i.i.d. (i.e., white noise). Note that, by increasing the order of the model (i.e., including more lagged terms), it is possible to remove the serial dependency of the residuals [25].

For instance, Fig. 1 depicts the autocorrelation plot of the residuals obtained with an AR model (that includes lags *t-1*, *t-2* and *t-24*) for one DTC. As depicted, the residuals are almost uncorrelated (i.e., in the sense that the autocorrelation values are inside the 95% confidence interval), which validates the choice of the lagged terms. This is the reason why the second lag was included in the model.

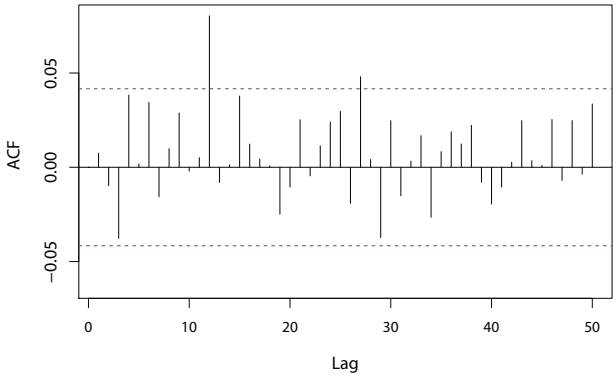


Fig. 1. Autocorrelation plot of the AR model's residuals.

Since the goal is to produce six hours-ahead forecasts, a different AR and VAR model is fitted for each lead-time. For instance, for lead-times 2 and 6, the VAR model has the following form:

$$P_{t+2} = \alpha_2 + \mathrm{B}_1 \cdot P_{t-1} + \mathrm{B}_2 \cdot P_{t-2} + \mathrm{B}_3 \cdot P_{t-22} + \mathrm{E}_{t+2} \quad (12)$$

$$P_{t+6} = \alpha_6 + \mathrm{B}_1 \cdot P_{t-1} + \mathrm{B}_2 \cdot P_{t-2} + \mathrm{B}_3 \cdot P_{t-18} + \mathrm{E}_{t+6} \quad (13)$$

where the terms $P_{t-1}$ and $P_{t-2}$ remain the same, and the seasonal effect associated to the previous day changes with the lead-time.

The RLS algorithm (Eq. 9-11) is used to update the coefficients of each lead-time (e.g., coefficients $B$ of Eq. 12 and 13).

Finally, in addition to the AR and VAR models described in the previous sections, a VAR with exogenous variables (VARX) is also proposed and tested. The model consists in adding exogenous variables to Eq. 7: the solar power values observed in each EB.

The goal is to assess if the EB measurements ($P_t^{EB}$) can improve the solar power forecast at the DTC level. The VARX for lead-time *t+1* has the structure depicted in Fig. 2. Note that, for the EB observations, only the lags *t-1* and *t-2* are included in the model since the goal is to use the EB as distributed sensors that characterize the current atmospheric conditions (in terms of solar power) across the region.

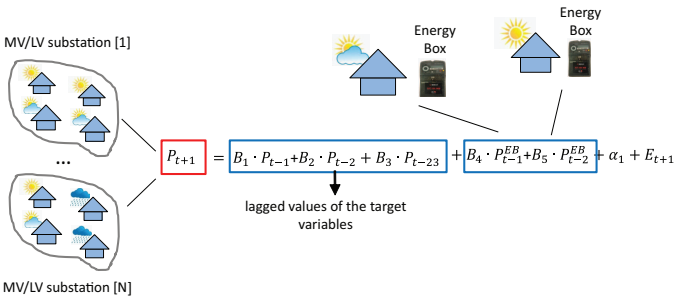


Fig. 2. VARX model for lead-time t+1 with EB measurements.

## IV. Test Case Results

### A. Description

The solar power dataset used as test case is from the city of Évora, with about 54,000 residents and an area of 1307 km$^2$. The distribution network in Évora municipality is supplied by two 60 kV substations with 15 kV and 30 kV feeders. 25 of these feeders supply a total of 655 secondary substations.

In what concerns distributed generation, in August 2013, there were 218 micro-generation producers, mainly solar PV, with an installed capacity of 761 kW. During 2 years, more than 30,000 EB and 300 DTC were installed in Évora, including all customers and substations, in order to have the entire municipality covered.

In order to test the proposed forecasting framework, time-series from 44 EB were used. These were the time-series with better quality, or in other words, the ones with the lowest number of missing values and hours with zero generation due to maintenance operation or communication problems. The EB data comprise domestic PV, with installed capacity ranging between 1.1 kWp and 3.7 kWp. These EB measurements were related to 10 different DTC, and the total values of each DTC are also forecasted.

The "optimal" parameters of the clear-sky model are $\sigma_h$=0.01, $\sigma_{doy}$=0.01, $\tau$=85%. The "optimal" *forgetting factor* $\lambda$ for both AR and VAR is equal to 0.999.

The original data was sampled in 15 minutes, but it was resampled to hourly values (i.e., the same length of the electricity market). The period between 1 February 2011 and 31 January 2012 was used to fit the models, and the period between 1 February 2012 and 6 March 2013 was used to calculate the forecast errors.

The forecasting results are evaluated with the root mean square error (RMSE) calculated for the $k^{th}$ horizon [10]:

$$\mathrm{RMSE}_k = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(\hat{P}_{t+k|t} - P_{t+k}\right)^2} \qquad (14)$$

The RMSE is normalized with the solar peak power.

The RMSE is calculated separately for each EB or DTC, but it is also calculated using the full dataset of errors as a summary performance metric for all DTC or EB. The performance of two models (AR and VAR) is compared by computing the improvement in terms of RMSE:

$$\mathrm{Imp}_k = \frac{RMSE_{k,AR} - RMSE_{k,VAR}}{RMSE_{k,AR}} \cdot 100\% \qquad (15)$$

### B. Results

The improvement of the VAR and VARX over the AR model for each lead-time is plotted in Fig. 3 for two DTC and for the RMSE calculated with the full dataset of DTC forecast errors.

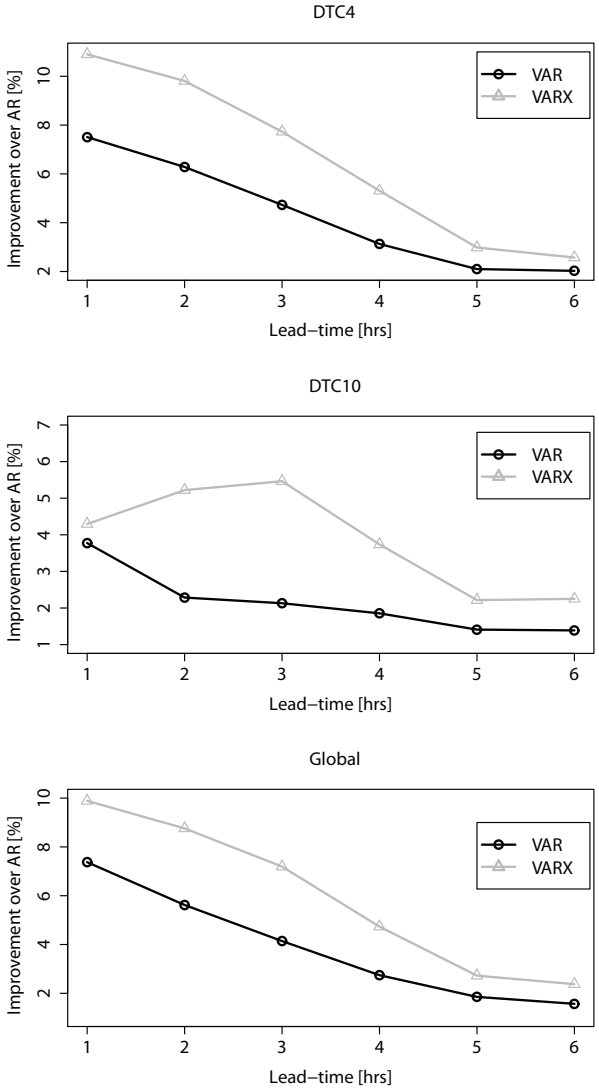


Fig. 3. $Imp_k$ of the VAR and VARX models for two DTC and for the RMSE calculated using the entire set of forecast errors.

These three plots clearly show that the VARX model achieves the highest improvement. From the full set of DTC, number 4 is the one with the highest overall improvement, reaching a value slightly above 10% for the first lead-time and around 2% for the sixth lead-time. Number 10 is the one with the lowest improvement, particularly between lead-times 1 and 4.

The VAR model also achieves a positive improvement in all lead-times, but lower than VARX. This means that the EB measurements, used as distributed sensors, can improve the forecast at the DTC level.

The global improvement varies between 10% and 2%, which shows the benefit from using spatial-temporal models for very short-term solar power forecasting. Another interesting conclusion is that the improvement decays with the lead-time, meaning that the spatial-temporal information is more relevant for the first three hours. This makes sense since the forecasting model in this test case only includes information from a small municipality. If solar power data from neighboring municipalities and regions is included in the model, a higher improvement for lead-times between 4 and 6 is expected.

Fig. 4 shows the improvement obtained with the VAR model for two EB and for the RMSE calculated with the full dataset of EB forecast errors.

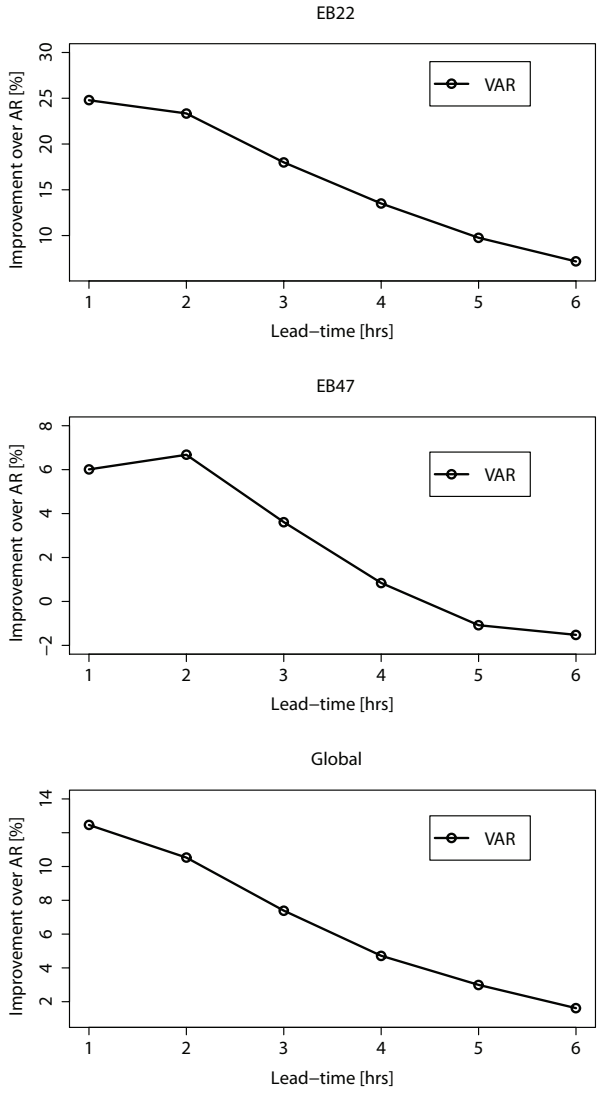


Fig. 4. $Imp_k$ of the VAR and VARX models for two EB and for the RMSE calculated using the entire set of forecast errors.

The VAR model attained the highest improvement for EB number 22, with a value of 25% for lead-time 1 and around 9% for lead-time 6. The lowest improvement was attained for EB number 47, with 6% for lead-time 1 and a negative value of around -2% for lead-time 6. The global improvement for the EB dataset varies between 12% and 3%. Compared to the DTC results, the improvement obtained for the EB dataset is higher.

Fig. 5 depicts the average, minimum and maximum values of the normalized RMSE for the EB and DTC datasets, calculated from the individual RMSEk values of each EB and DTC. For the DTC, the forecast errors are from the VARX model, while for the EB are from the VAR model.

The average RMSEk of the EB and DTC is similar, but marginally higher for the EB. The main difference is in the minimum and maximum values, with a higher amplitude for the EB. For instance, there are EB with an RMSE below 10% (even for lead-time 6), but also EB with an RMSE close to 20%. This is an expected result, since the variability in solar power due to the clouds is smoothed by the aggregation of EB in one DTC. Nevertheless, this also shows that the forecasts for some EB are significantly improved with the spatial-temporal information (e.g., EB number 22 in Fig. 4).

Fig. 6 depicts the improvement of the VAR over the AR model, calculated using the entire set of EB forecast errors and for three seasons (winter, spring and summer).

The highest improvement for the first three lead-times occurs in the winter season, which frequently exhibits overcast days (“stable” clouds). The summer season frequently has clear-sky days, thus the improvement of the VAR model is lower. Nevertheless, for the last three lead-times, the results are contrasting and the improvement is higher in the summer season. This might be explained by the presence of small and rapid clouds (“unstable” clouds) during summer, which are captured by the six hours time-window. However, additional weather data (e.g., from weather stations) that would enable a more detailed analysis are needed to explain these differences.

Finally, it is important to emphasize that the VAR model showed positive improvement in all three seasons.

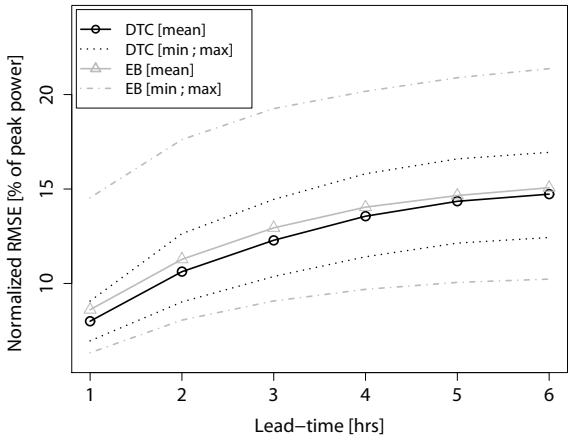


Fig. 5. Average, minimum and maximum normalized $RMSE_k$ calculated with the individual $RMSE_k$ obtained for each EB and DTC.

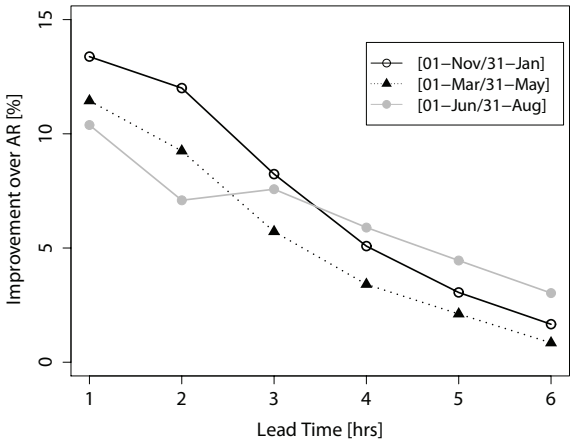


Fig. 6. Improvement of the VAR over AR for different seasons of the year, calculated with the full set of forecast errors for each EB.

## V. CONCLUSIONS

A new forecasting approach for very short-term solar power forecast, based on a vector autoregressive model fitted with recursive least squares, is proposed in this paper. It takes advantage of a Smart Grid infrastructure with smart meters and advanced control functions installed at the MV/LV substation level.

The results for data from a Smart Grid pilot, in the city of Évora, Portugal, indicate that information from distributed PV generation can improve the forecast error, compared to an autoregressive model, between 12% and 8% on average for the first 3 lead-times. Furthermore, the inclusion of the EB observations in the forecast at the DTC level also decreases the error. Therefore, the adoption of multivariate models such as VAR with an online tracking of its coefficients is recommended to improve the solar power forecasting in a Smart Grid environment. This forecasting framework can be explored either by the DSO or by a solar PV aggregator.

These results open new lines for future research and improvement, such as: development of a feature selection algorithm that reduces the amount of information used and consequently the communication requirements; inclusion of data from weather stations and numerical weather prediction models; increasing the spatial coverage of the PV data.

## REFERENCES

[1] P. Hummel, P. Lekander, A. Gandolfi, S. Hunt and I. Cossio, "The unsubsidised solar revolution," Technical Report, UBS Investment Research, January 2013.

[2] H. Farhangi, "The path of the Smart Grid," *IEEE Power and Energy Magazine*, vol. 8, no. 1, pp. 18-28, 2010.

[3] A. Madureira, L. Seca, J. Peças Lopes, P. Matos, N. Silva, "Maximizing the integration of distributed generation in smart grids distribution systems," in *Proceedings of the CIGRE Symposium*, Lisbon, Portugal, April 2013.

[4] A.G. Madureira, J.A. Peças Lopes, "Ancillary services market framework for voltage control in distribution networks with microgrids," *Electric Power Systems Research*, vol.86, pp.1-7, May 2012.

[5] J. Krstulovic, V. Miranda, A. Simões Costa, J. Correia Pereira, "Towards an auto-associative topology state estimator," *IEEE Transactions on Power Systems*, vol.28, no.3, pp.3311-3318, August 2013.

[6] J. Soares, J.A. Peças Lopes, "Controlling electric vehicles in quasi-real-time," in *Proceedings of the IEEE PowerTech 2013*, Grenoble, France, 16-20 June 2013.

[7] European Photovoltaic Industry Association (EPIA), "EPIA Response to DG ENER working paper - The future role and challenges of energy storage," Position Paper, February 2013.

[8] Website (accessed on November 2013): http://www.renewableenergyworld.com/rea/news/article/2013/03/solar-storage-market-set-for-rapid-growth?cmpid=rss

[9] C. Monteiro, R. Bessa, V. Miranda, A. Botterud, J. Wang, G. Conzelmann, "Wind power forecasting: state-of-the-art 2009," Report ANL/DIS-10-1, Ar-gonne National Laboratory, November 2009.

[10] P. Bacher, H. Madsen, H.A. Nielsen, "Online short-term solar power forecasting", *Solar Energy*, vol. 83, no. 10, pp. 1772-1783, October 2009.

[11] L.A. Fernandez-Jimenez, A. Muñoz-Jimenez, A. Falces, M. Mendoza-Villena, E. Garcia-Garrido, P.M. Lara-Santillan, E. Zorzano-Alba, P.J. Zorzano-Santamaria, "Short-term power forecasting system for photovoltaic plants," *Renewable Energy*, vol. 44, pp. 311-317, August 2012.

[12] A. Hammer, D. Heinemann, C. Hoyer, R. Kuhlemann, E. Lorenz, R. Müller, H.G. Beyer, "Solar energy assessment using remote sensing technologies," *Remote Sensing of Environment*, vol. 86, no. 3, pp. 423-432, August 2003.

[13] H. Pedro, C. Coimbra, "Assessment of Forecasting Techniques for Solar Power Production with no Exogenous inputs," *Solar Energy*, vol. 86, no. 7, pp. 2017-2028, July 2012.

[14] H.M. Diagne, M. David, P. Lauret, J. Boland, "Solar irradiation forecasting: state-of-the-art and proposition for future developments for small-scale insular grids," *World Renewable Energy Forum (WREF 2012)*, Colorado, USA, 13-17 May 2012.

[15] V. Berdugo, C. Chaussin, L. Dubus, G. Hebrail, V. Leboucher, "Analog method for collaborative very-short-term forecasting of power generation from photovoltaic systems," *Next Generation Data Mining Summit (NGDM '11)*, Athens, Greece, 4 September 2011.

[16] C. Yang, L. Xie, "A novel ARX-based multi-scale spatiotemporal solar power forecast model," in *Processings of the North American Power Symposium (NAPS)*, USA, September 2012.

[17] P. Godinho Matos, P. Daniel, A. Veiga, A. Messias, M. Oliveira, P. Monteiro, "InovGrid, a smart vision for a next generation distribution system," in *Proceedings of the 22nd International Conference on Electricity Distribution (CIRED)*, Stockholm, Sweden, 10-13 June 2013.

[18] P. Lúcio, P. Paulo, H. Craveiro, "InovCity - Building smart grids in Portugal," in *Proceedings of the 21nd International Conference on Electricity Distribution (CIRED)*, Frankfurt, Germany, 6-9 June 2011.

[19] X. Lu, W. Wang, J. Ma, "An empirical study of communication infrastructures towards the smart grid: design, implementation, and evaluation," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 170-183, March 2013.

[20] P. Bacher, "Short-term solar power forecasting," MSc Thesis, Technical University of Denmark, 2008.

[21] R. Koenker, G. Bassett, "Regression quantiles", *Econometrica*, vol. 46, pp. 33-50, 1978.

[22] H. Madsen, *Time Series Analysis*, London: Chapman and Hall/CRC, 2006.

[23] R. Davidson, J.G. MacKinnon, *Econometric Theory and Methods*, New York: Oxford University Press, 2003.

[24] L. Ljung, T. Soderstrom, T*heory and Practice of Recursive Identification*, Cambridge: The MIT Press, 1983.

[25] J.M. Wooldridge, *Introductory Econometrics: A Modern Approach*, 2nd Edition, South-Western College Pub, 2002.